THE UNIVERSITY
of EDINBURGH

The genetic underpinnings of adiposity: from GWAS discovery to functional characterisation

Katherine A Kentistou, *BSc, MSc, DIC*

Doctor of Philosophy

The University of Edinburgh

2020

Declaration

a) this thesis has been composed by me

b) the work is my own except as indicated below

c) the work has not been submitted for any other degree or professional qualification

d) any included publications are my own work, except where indicated below

Katherine A Kentistou, 2020

Abstract

Obesity affects more than 600 million people worldwide and causes 4 million excess deaths every year through its related co-morbidities, such as type 2 diabetes and heart disease. Detrimental anatomical distribution of body fat, specifically visceral fat accumulation, is a major driver of this mortality risk. There is clear evidence that underlying genetic predisposition plays a major role in determining body fat mass and distribution. Most genome-wide association studies to date have focused on anthropometric, rather than direct, measures of adiposity, such as BMI and WHR, and have successfully identified over a thousand candidate loci. However, the mechanisms via which these loci exert their effect is most often unknown. Thus, our understanding of the full impact and mechanisms underpinning the genetic contribution to human adiposity is incomplete. Here, we aim to further understand how human genomic variation contributes to body composition by using refined measures of adiposity and by gaining functional insight into the role of the resulting candidate loci.

We worked with non-invasive imaging phenotypes in UKBiobank, a population cohort that includes 500,000 participants from across the UK. Of these, 5,000 have undergone DXA scans that allow for the direct measurement of fat and lean mass throughout the body. We first established a way to impute the DXA adiposity patterns in the rest of the UKBiobank cohort, using anthropometric measures of body size. These imputed phenotypes provided accurate predictions of the underlying adiposity traits (R~0.8) and were used to increase our novel discovery power in a genome-wide association framework. Out of the hundreds of

associated loci, we sought to replicate the 27 strongest associations (P-value<$6x10^{-13}$) in other population cohorts for the underlying DXA measures. To do so, we worked with genetic and DXA data from the ORCADES, EPIC-Norfolk and Fenland studies (collective DXA n=18,000). Six of the submitted SNPs were replicated successfully (FDR 10%), while the vast majority (19/27) were directionally consistent between discovery and replication.

Two biologically promising candidate loci were followed-up to establish causal genes. The first locus, around gene *PLA2G6*, had previously been associated with various metabolic and obesity measures. However, computational approaches indicated several likely causal genes at the locus. We therefore used chromatin conformation capture technologies to pinpoint which of these genes interact with the associated interval. In doing so, we observed long-range interactions with three other genes *Pick1*, *Dmc1* and *Sox10*. The second locus was composed of eQTLs that changed the expression of gene *ADAMTS14* and was thus, further investigated it in a mouse model to ascertain the effect of *Adamts14*-null mutations on body composition in an *in vivo* setting. In the GWAS data, we saw a reduction of 18g in leg fat mass per copy of the effect allele and also a reduction in the expression of the gene. The *Adamts14*$^{-/-}$ mouse phenocopied this effect, as it showed reduced adiposity and weight gain under obesogenic conditions. It also showed increased energy expenditure and food intake, indicating an altered energy homeostasis.

Large scale biobanks offer great gene discovery potential. Here, we harnessed the phenotypic breadth of UKBiobank to impute direct body composition measurements and identify novel loci that may affect adiposity. Using

one such locus, we successfully replicated the human adiposity phenotype in a mouse model, thus validating this approach and highlighting its potential to offer biological insights into the genetics of obesity.

Lay summary

Obesity is a serious health issue and happens when we consume more calories than we burn, over long periods of time. Our bodies store these extra calories in fat all over our body, but if there is too much fat around our vital organs, their function can be affected. This type of body shape is often called apple shape and it has been shown to increase the chances of developing health problems, such as heart disease and type 2 diabetes. This is important because 4 million deaths every year are caused by obesity-related disease. However, other body shapes, such as pear shape, when fat is stored on the lower half of the body, do not carry the same disease risk.

I am interested in how our DNA, the building block of life, plays a role in developing these diseases. Our DNA is inherited through the generations and it can determine the way we carry fat on our bodies. DNA is organised into units called genes, each of these is responsible for producing a component of our bodies. When DNA gets copied, to be passed down from parent to child, mistakes often happen. These changes are usually harmless and have no effect on health, but sometimes they can affect the way our bodies work. Some can affect our body shape and these are the DNA changes I have been researching. If we could predict

who is more likely to have an apple shape based on their DNA, then we could protect the people who have the highest risk of developing disease.

To do this I have been working with data from participants all across the UK, who have undergone whole body scans, called DXA scans. These scans tell us exactly how much and where fat is on the body. Using this information, alongside the DNA sequence for all these participants, we can do simple tests of association between the presence of DNA changes and the amount of harmful fat. We did this for millions of mutations across the DNA and discovered many new genes that affect where fat accumulates in our body. But we still did not know how they do so, or how to intervene to minimise disease risk. To figure this out, we studied a special type of mouse that is missing one of these genes, called *Adamts14*. In our human dataset, people with mutations in *Adamts14*, produced this gene in less quantity and also had less fat in their legs. The mice missing *Adamts14* were quite similar, they weighed less and had less fat. We also looked at how much food they consumed and found that even though they are thinner, they are eating more food. This tells us that the gene *Adamts14* not only affects how the body develops, but also affects how the body burns calories. More work is needed to understand exactly how this happens.

To summarise, our DNA affects our risk of being obese and of developing disease. Understanding how different body shapes affect this risk can lead to new ways of preventing disease. I identified many new genes that affect body shape and studied a mouse that is missing the gene *Adamts14*. This work has been the first step to understanding exactly what these genes do in our bodies and how we can use them to prevent obesity and related diseases.

Acknowledgements

Firstly, I would like to thank the MRC Doctoral Training Programme in Precision Medicine and the University of Edinburgh for funding this work. Secondly, my supervisors, Nik Morton, Peter Joshi and Jim Wilson for supplying me with guidance, data and equipment throughout this project and my thesis committee, Philippa Saunders and Evropi Theodoratou, for their insights and continued help and support.

Thanks to all the study participants of the ORCADES and UK Biobank cohorts and everyone who contributed to the creation, maintenance and handling of this data. To Zoltan Kutalik for suggesting the DXA imputation work to me and hosting us at frosty Lausanne during the cold wave of 2018. To Wendy Bickmore and Iain Williamson for their enthusiastic contribution to this project. To all the staff at LF2 especially Will, Sandra, Laraine and Hollie for their invaluable support to projects like this.

I would like to thank everyone in the Wilson and Molecular Metabolism groups, DTP student cohorts and Adipose Biology Club, for the sharing of skills and knowledge. But especially, Paul T and Erin MD, my QuantGen travel buddies, for late night genetics chats around the globe. Joanna S and Jordan P for making the beginning of my PhD as painless as possible. Rod C for his boundless lab wisdom and patience. Tanguy B for scruffing encouragement. Loes E, Ineke L and Eleanor M who made every day slightly brighter. Marlene MP for the endless supply of biscuits. Peter J for sharing some of his scallops with me towards the end. I feel incredibly grateful to Jim D, my partner in thesis write-up, lockdown and general life, who kept me sane with good music, good food and endless laughter. Finally, I would like to thank my family and especially my mum, for their sacrifices which have allowed me to be where I am today.

Ευχαριστώ!

Assistance

Drafting | My supervisors commented on chapter drafts with regards to grammar, rewordings, additional references and required changes to figures and tables.

Human Cohort Data | The phenotypes, called array and imputed genotypes were provided to me for subsequent analysis; in the case of ORCADES by Jim Wilson and Peter Joshi and for UKB by David Clark via the UKB website.

GWAS | GWAS in ORCADES and UKB were conducted using a pipeline created by Peter Joshi, David Clark, Paul Timmers and Andrew Bretherick.

Replication | Replication in the two independent cohorts, EPIC-Norfolk and Fenland, was carried out by members of the Langenberg Group, Laura Wittemans and Jian'an Luan.

TWMR | This analysis was carried out by Eleonora Porcu.

3C & 5C | This experiment was carried out in collaboration with the Bickmore Group. Iain Williamson taught and supervised me while doing this experiment.

Mouse Experiments | I received training for oral gavages and tail nicks by William Mungall. Daily care and maintenance of the mouse colonies were carried by the staff of the BRR unit at the Edinburgh Bioquarter.

Histology | Post mortem tissues were sectioned by members of the CALM facility.

Contribution to publications

Kentistou, K. A., Wilson, J. F., Joshi, P. K., & Morton, N. M. (2019). The genetic underpinnings of obesity. Current Opinion in Physiology, 12, 57-64.

Joshi, P. K., Pirastu, N., Kentistou, K. A., Fischer, K., Hofer, E., Schraut, K. E., et al. (2017). Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. Nature communications, 8(1), 910.

Shrine, N., Guyatt, A. L., Erzurumluoglu, A. M., Jackson, V. E., Hobbs, B. D., Melbourne, C. A., et al. (2019). New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. Nature genetics, 51(3), 481-493.

Evangelou, E., Gao, H., Chu, C., Ntritsos, G., Blakeley, P., Butts, A.R., Pazoki, R., Suzuki, H., Koskeridis, F., Yiorkas, A.M. and Karaman, I., et al. (2019). New alcohol-related genes suggest shared genetic mechanisms with neuropsychiatric disorders. Nature human behaviour, 3(9), pp.950-961.

Pirastu, N., McDonnell, C., Grzeszkowiak, E. J., Mounier, N., Imamura, F., Day, F. R., et al. (2019). Using genetics to disentangle the complex relationship between food choices and health status. bioRxiv, 829952.

Conference abstracts

American Society of Human Genetics meeting, October 2018, Orlando, Florida, USA (Poster presentation) "Analysis of 500,000 lives reveals gene-environment interactions for social class and obesity"

European Mathematical Genetics Meeting, April 2018, Cagliari, Sardinia, Italy (Poster presentation) &

Centre for Cardiovascular Science annual symposium, June 2018, Edinburgh, UK (Poster presentation) "Assessing novel anthropometric indices and their predictive efficacy over metabolic health and disease"

Orkney Science Festival, September 2019, Kirkwall, Orkney, UK (Oral presentation) "How big are your genes?"

American Society of Human Genetics meeting, October 2019, Houston, Texas, USA (Poster presentation) "Mendelian randomisation reveals that testosterone causally improves the metabolic profile of middle-aged men"

List of Abbreviations

3C	chromatin conformation capture
3D-FISH	3D fluorescent in situ hybridisation
5C	3C carbon copy
ABSI	a body shape index
AT	adipose tissue
BAI	body adiposity index
BAT	brown AT
BIA	bioelectrical impedance analysis
BMI	body mass index
BRI	body roundness index
CNS	central nervous system
CT	computerised tomography
CVD	cardiovascular disease
DXA	dual-emission x-ray absorptiometry
E(1-21)	embryonic day (1-21)
ECM	extracellular matrix
EE	energy expenditure
EPIC	European prospective investigation of cancer
eQTL	expression quantitative trait loci
FBS	foetal bovine serum
FDR	false discovery rate
FMI	fat mass index

FMR	fat-to-lean mass ratio
GF	germ free
GFP	green fluorescent protein
GWAMA	genome wide association meta-analysis
GWAS	genome wide association study
GWS	genome wide significant/significance
HEIDI	heterogeneity in independent instruments
HFD	high fat diet
HWE	Hardy-Weinberg equilibrium
iDXA	imputed DXA
IVF	in vitro fertilisation
KO	knock out
LD	linkage disequilibrium
LDL	low density lipoprotein
LMI	lean mass index
MA	meta-analysis
MAF	minor allele frequency
MR	Mendelian randomisation
MRI	magnetic resonance imaging
MSC	mesenchymal stem cell
oGTT	oral glucose tolerance test
OR	odds ratio
P(1-21)	postnatal day (1-21)
PBS	phosphate buffered saline

PC	principal component
PFA	paraformaldehyde
PSR	picrosirius red
QC	quality control
RER	respiratory exchange rate
RT	room temperature
SAT	subcutaneous AT
SES	socioeconomic status
SFI	segmental fat index
SMR	summary-data-based Mendelian randomisation
SNP	single nucleotide polymorphism
T2D	type 2 diabetes
TD-NMR	time domain nuclear magnetic resonance
TFI	trunk fat index
TLI	trunk lean index
TWMR	transcriptome-wide summary statistics-based MR
UKB	UK biobank
VAT	visceral AT
WAT	white AT
WC	waist circumference
WHO	world health organisation
WHR	waist-hip ratio
WHRadjBMI	WHR adjusted for BMI
WT	wild-type

List of Genes

ACADVL	acyl-coa dehydrogenase very long chain
ACVR2B	activin a receptor type 2b
ADAM	a disintegrin and metalloproteinase domain (gene family)
ADAMTS	*ADAM* with thrombospondin type 1 (gene family)
CD300a	*CD300a* molecule
CPS1	carbamoyl-phosphate synthase 1
CTCF	*CCCTC* binding factor
DLG4	discs large maguk scaffold protein 4
DMC1	dna meiotic recombinase 1
EXOG	exo/endonuclease g
FBXO36	f-box protein 36
FTO	fat mass and obesity associated
IL1b	interleukin 1 beta
IL6	interleukin 6
IRX3/5	iroquois homeobox 3/5
MAFF	maf bzip transcription factor f
MC4R	melanocortin 4 receptor
MMP	matrix metalloproteinase (gene family)
PICK1	protein interacting with prkca 1
PLA2G6	phospholipase a2 group vi
RPGRIP1L	rpgrip1 like
RSPO3	r-spondin 3
SORT1	sortilin 1
SOX10	sry-box transcription factor 10
TMEM18	transmembrane protein 18
TNFa	tumour necrosis factor-alpha
TRIP12	thyroid hormone receptor interactor 12
UCP1/3	uncoupling protein 1/3

Table of Contents

List of Figures

List of Tables

1

Introduction

1.1 Obesity as a public health concern

Obesity is the fifth leading cause of death and affects more than 600 million adults and 100 million children worldwide (GBD, 2017). It is commonly defined as a body mass index (BMI) over 30 kg/m^2 and is most prevalent among the middle-aged population. In 2015, high BMI (>25 kg/m^2) contributed to 4 million deaths and 120 million disability-adjusted life years worldwide, via its many comorbities (GBD, 2017).

The rise in the prevalence of obesity happened almost simultaneously in most high-income countries in the 1970s and 1980s (Sassi et al., 2009). The observed paradigm across many countries was that groups of high socioeconomic status (SES) in urban areas of low or middle-income countries were the first to be affected. This shifted to groups of low SES with a rise in the country's gross national product (Monteiro et al., 2004). Throughout this transition, prevalence was always highest in women and middle-aged individuals (Bourne and

Collaborators, 2016). As of 2017, the global prevelance of obesity in middle-aged males and females rose from 7% to 12% and 13% to 20%, respectively, since 1980. The prevelance of overweight in the same groups went from 32% to 45% in men and 38% to 49% in women (IHME, 2015, GBD, 2017).

Obesity poses a singular threat to public health for two reasons. Firstly, obesity is one of the only leading health risks that continues to increase in prevalence. The rates of increase are largely comparable between different countries and socioeconomic strata, which suggests that rates of increase are not a direct consequence of affluence (Bourne et al., 2016). Changes in the food environment are also likely drivers of this surge. In recent years, there has been a rise in the availability and affordability of processed, highly energy-dense foods, which facilitate excess energy intake. Arguably, the urbanisation of modern societies with all of its conveniences, brought along a decrease in manual labour and generalised physical activity which contributed to the reduced expenditure of energy.

Secondly, despite several attempts, there are no examples in which the obesity epidemic has been reversed by interventions, as has been the case for many other preventable death causes (i.e. smoking, infectious disease, etc.)

(Swinburn et al., 2011). Obesity is often attributed to a complex interplay between intrinsic genetic susceptibility and modern obesogenic environments (Barness et al., 2007), thus it could be seen as a biological response to these emerging environments. However, to deliver effective interventions, priority should be given to addressing the environments that facilitated the problem in the first place. Namely, policies and priorities in modern societies promote overconsumption as a model for economic growth; obesity is only one of the consequent crises of this phenomenon. Any effective solutions will need to be government-led and education-based and prioritise sustainable and systemic long-term changes, directed at societal environments and not at the individual (Swinburn et al., 2011).

1.2 Obesity disease sequelae & body fat distribution

Such interventions are of paramount importance as obesity is a well-established risk factor for cardiovascular disease (CVD), type 2 diabetes (T2D) and various cancers (WHO, 2011). In 2017, high BMI contributed to 3 million deaths attributable to CVD, 760,000 deaths from T2D complications and 460,000 deaths due to cancer (IHME, 2015). This brought the global death rate due to high BMI up by almost 30%, across the queried 27-year period.

Although, obesity is simply defined as a BMI>30, the presentation of the condition is heterogeneous. Adipose tissue (AT) can be distributed in many different ways throughout the human body and is often categorised into two main patterns (Figure 1.1). The first type of adipose distribution is central obesity, whereby fat accumulates in the upper half of one's body and particularly when excess visceral adipose tissue (VAT) exists within the abdominal cavity and surrounds the intra-abdominal organs. This distribution pattern greatly increases the risk of CVD and T2D (Liu et al., 2010). Contrarily, lower-body adiposity and particularly gluteofemoral subcutaneous adipose tissue (SAT), reduces risk (Figure 1.1) (Shuster et al., 2012, Karpe and Pinnick, 2015).

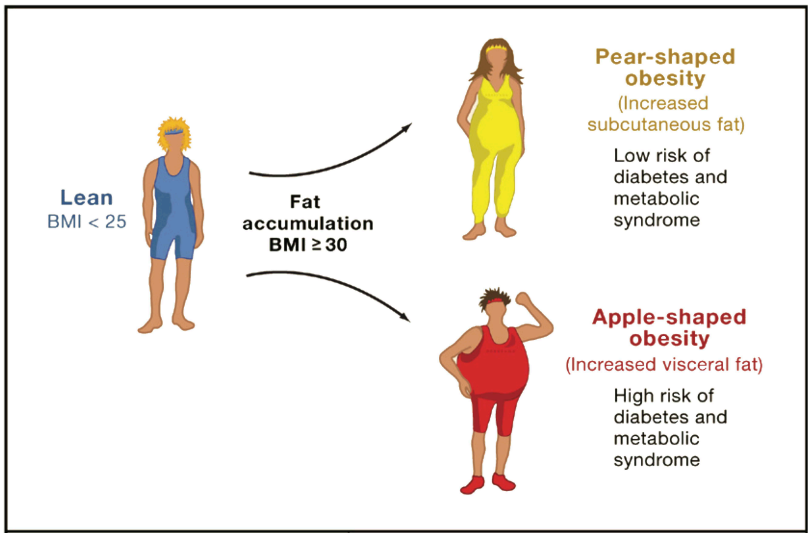


Figure 1.1 | Body shape influences metabolic disease risk.

Obesity is often defined as having a BMI>30. However, apple-shaped individuals more intra-abdominal visceral fat and increased chances of developing cardiometabolic complications, while pear-shaped individuals have a lower risk of developing disease. *Adapted from (Gesta et al., 2007).*

It is also worth noting, that in certain situations reduced adiposity can confer an increased disease risk. Namely, the abnormal paucity of AT, characteristic of several lipodystrophic conditions carries the same clinical complications as excess AT; insulin resistance and T2D, fatty liver, dyslipidaemia and hypertension (Savage et al., 2003, Kim et al., 2008). In many of these lipodystrophic cases and as seen previously with apple shaped individuals, this propensity to T2D, CVD and other metabolic dysfunctions has been shown to be caused specifically by the absence of SAT (Savage et al., 2003, Scott et al., 2014, Yaghootkar et al., 2014).

These AT distribution patterns are highly age and sex dependent. Specifically, women accumulate more SAT peripherally (pear-shape), compared to men (Pasco et al., 2014). This distinction is maintained until menopause, when both sexes accumulate more VAT (apple-shape) (Wells, 2007). Therefore, to truly understand and prevent the effects of fat distribution on human disease, one must carefully consider its different aetiological origins and also the way obesity is currently assessed.

1.3 Ways of assessing obesity & adiposity

The World Health Organization (WHO) defines obesity as the condition when fat accumulates abnormally or excessively in adipose tissue, to an extent that may cause disease (WHO, 2000). Assessing fat accumulation ideally requires direct measurement of the fat depots. However, due to the expense and logistical barriers associated with accurate non-invasive imaging methodologies for this, there is a number of simple measures that are commonly used to infer adiposity and obesity.

1.3.1 Anthropometric measures

Anthropometry is a simple non-invasive way to infer body shape. Such measures are easy to attain and calculate, but are often crude representations of the underlying body composition. The most commonly used anthropometric measure to define obesity is BMI, which can be calculated by dividing one's weight in kilograms by height in meters squared (Table 1.1) (NIH, 1998). BMI ranges are used to classify individuals into underweight (<18.5), normal weight (18.5-24.99), overweight (25-29.99) and obese (>30). However, BMI cannot account for different body shapes and compositions. For example, it cannot distinguish

between lean and fat mass. Therefore, while somewhat useful at a population scale, it can often misclassify individuals.

An alternative to BMI is waist-hip ratio (WHR). WHR is calculated by dividing one's waist circumference by their hip circumference (Table 1.1) and is thus more indicative of central adiposity. This has been corroborated by an increased risk of disease. Namely, it has been shown that the relative risk of myocardial infarction is 0.10 and 1.37, while the risk of T2D is 1.87 and 1.88, per standard deviation of BMI and WHR, accordingly (Vazquez et al., 2007, Yusuf et al., 2005). Conventionally, WHR>1.0 in men and >0.85 in women serves as an indicator of excess abdominal fat accumulation (WHO, 2011). Waist circumference (WC) alone can also be used to the same effect, but without taking gluteofemoral fat into account, which WHR might be able to infer. WC of more than 94 cm in men and 80 cm in women indicates an increased risk of metabolic complications (WHO, 2000).

Notably, BMI becomes less predictive of metabolic dysfunction in elderly individuals (Stevens et al., 1998), while other anthropometric measurements such as WC and WHR remain predictive in old age. However, WHR and WC appear to be highly population specific. Namely, it has been shown that at a given WC, Asian

populations have more VAT and African populations have less, when compared to Europeans (WHO, 2011). Thus, care should be taken when using the above thresholds.

Effort has also gone into developing more informative indices that combine several anthropometric measurements in order to create more accurate predictions of central adiposity. One such measure is "a body shape index" (ABSI), which combines WC, BMI and height (Table 1.1) and is reportedly a better predictor of central adiposity and premature mortality than any of its components (Krakauer and Krakauer, 2012). A similar measure, body roundness index (BRI) is based on an individual's physical eccentricity, i.e. their non-dimensional geometric value, which estimates the size of the human body as an ellipse and is inferred by their WC and height (Table 1.1). BRI has been similarly utilised and is predictive of body adiposity as well as generalised metabolic health (Thomas et al., 2013). Body adiposity index (BAI) can be calculated from hip circumference and height (Table 1.1), and showed a higher correlation with body fat, than BMI (Bergman et al., 2011).

Table 1.1 | Formulae used to calculate composite phenotypes and indices.

Weight in kg, height, waist and hip circumference in cm.

MEASURE	CALCULATION
ABSI	$ABSI = \frac{waist\ circumference}{BMI^{2/3} \times height^{1/2}}$
BAI	$BAI = \frac{hip\ circumference}{height^{3/2}} - 18$
BMI	$BMI = \frac{weight}{height^2}$
BRI	$BRI = 364.2 - 365.5 \times \sqrt{1 - \left(\frac{\frac{waist\ circumference}{2\pi}}{0.5 \times height}\right)^2}$
WHR	$WHR = \frac{waist\ circumference}{hip\ circumference}$

With several studies supporting or refuting the superior discriminative capacities of these measures for detrimental body fat distribution patterns (López et al., 2012, Bennasar-Veny et al., 2013, Lichtash et al., 2013, He and Chen, 2013, Duncan et al., 2013, Maessen et al., 2014), it is often unclear whether they remain equally informative for individuals across all age-ranges, sexes and ethnic backgrounds.

1.3.2 Bioelectrical impedance analysis (BIA)

BIA is another widely used method for inferring VAT, that is low-cost and has often been used in a clinical setting (Kyle et al., 2004). BIA conducts an

electrical current through the body and measures the resistance encountered by the fat and lean parts of the body. It estimates whole body fat content with some accuracy, but care should be taken when trying to calculate VAT and SAT via BIA, as these are merely inferred with this technology (Nagai et al., 2008).

1.3.3 Direct imaging of body composition

In contrast to anthropometric and BIA measures, imaging platforms can provide more precise measurements of VAT mass and other adipose depots, as outlined below.

1.3.3.1 Dual-emission X-ray Absorptiometry (DXA)

DXA is a rapid and precise whole-body imaging technology. Unlike other technologies, DXA scans offer the possibility of whole-body, as well as regional analyses (Mourtzakis et al., 2008). However, DXA is only able to infer VAT given its location, rather than directly measure it. It measures the attenuation of two energies to distinguish the fat, lean and bone contents of the body based on their elemental differences (Silver et al., 2010). It does so by separating the body into distinct predetermined areas, as seen in Figure 1.2, and characterising their composition. Although DXA sometimes underestimates fat mass, particularly in

cases of extreme obesity, it remains a far better tool at inferring body shape than BIA or anthropometry and is much more cost effective than other imaging platforms (Levitt et al., 2010, Ball and Swan, 2003).

1.3.3.2 Magnetic resonance imaging (MRI) & computerised tomography (CT)

While DXA scans are quite efficient at inferring adiposity patterns, MRI and CT scans are currently considered the gold standard for quantifying abdominal

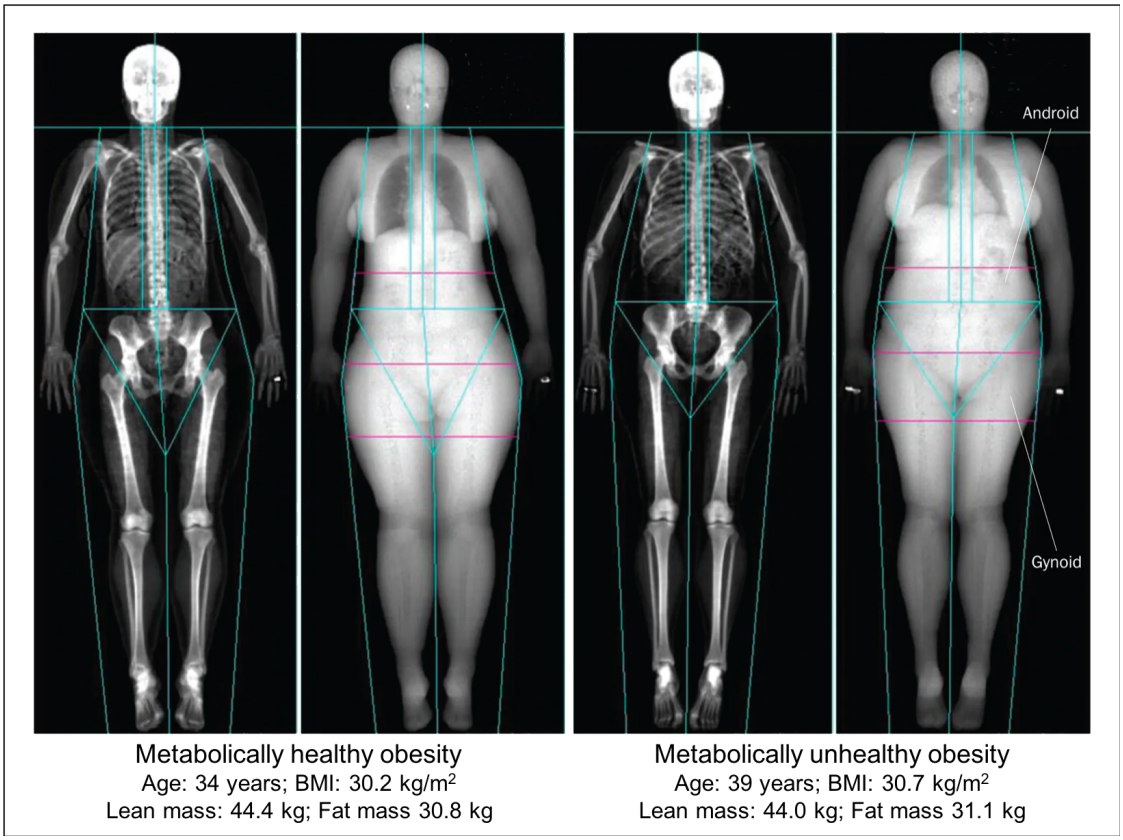


Figure 1.2 | Body composition measured by DXA.

Two female individuals matched for age, lean and total fat mass, but exhibit very distinct body shapes. The woman on the left exhibits a lower-body/gynoid/pear-shaped pattern while the one on the right shows a central/android/apple-shaped pattern of adipose tissue distribution. *Modified from (Karpe and Pinnick, 2015).*

adipose tissue. Contrarily to DXA scans, they both work with cross-sectional images and can directly measure the volume of VAT and SAT within each queried region. In clinical practice, "single-slice" images at specific pre-determined landmarks are often used to reduce radiation exposure and costs, but these are often inaccurate, given the mobility of the queried tissues (Greenfield et al., 2002, Thomas et al., 1998). Additionally, CT and MRI image analysis often requires more time and expertise than DXA.

1.4 The biological basis of obesity

While excess adiposity confers added health risks, adipose tissue is essential for the survival of most species. All eukaryotic organisms have the capacity to store calories in the form of lipid droplets and all vertebrates do so in specialised cells that are called adipocytes (Ottaviani et al., 2011). Adipocytes are critical for survival and responsible predominantly for catering to an organism's energy and nutritional homeostasis, but also serve other purposes, such as immunologically and mechanically protecting the other organs, providing insulation and transducing signals through the body (Tchkonia et al., 2013, Thomou et al., 2010, Duffaut et al., 2009).

Adipocytes aggregate and develop in many depots throughout the human body. They exist in two broad types of AT; white (WAT) and brown (BAT). BAT has the ability to dissipate stored energy in the form of heat, while WAT retains it in triglycerides. In adult humans, BAT is often referred to as beige AT, due to its hybrid nature and distinct cellular origin, compared to rodent BAT (Kajimura et al., 2015). WAT depots are further separated into visceral or subcutaneous, the latter existing just under the epidermis and the former within the body cavity. Each of these depots has distinct biological signatures and reacts differently towards metabolic challenges.

In the case of obesity, energy intake exceeds energy expenditure, white adipocytes expand and replicate and when the ability of adipocytes to store excess energy is exceeded, lipids get stored in non-adipose tissues, causing adverse effects on normal body function, as outlined below.

1.4.1 Development and normal function of AT

In humans, adipose tissue develops during the second trimester of gestation and WAT is apparent by the 14th week (Poissonnet et al., 1984, Pellegrinelli et al., 2016). Adipocyte proliferation slows down towards the end of gestation and up until puberty AT mass increases mainly by the filling of pre-

existing adipocytes. Puberty is characterised by a further increase in AT cellularity, after which the total number of adipocytes is set within an individual (Knittle et al., 1979). In mice, the development of SAT precedes that of VAT. SAT forms between E14 and 18 (Wang et al., 2013) and expression of adipocyte markers is detectable in the subcutaneous region at E16.5–17.6 (Birsoy et al., 2011). VAT forms postnatally and precursors can be found in the epididymal pad at P4, while they only become visible by P7 (Han et al., 2011).

Adipocytes are of a mesodermal cellular origin. Mesenchymal stem cells (MSCs) in the mesoderm give rise to WAT precursors, which in turn give rise to adipocytes (Billon et al., 2007). MSCs are also capable differentiating into osteoblasts, myoblasts and connective tissues. The number of intermediate stages separating MSCs from mature adipocytes remains unclear, but they can also give rise to committed white or brown AT precursors (Timmons et al., 2007, Seale et al., 2008, Lepper and Fan, 2010).

Once developed, AT has many functions and interacts with many other types of cells within the body. Adipocytes only make up 20–40% of the cellular content of a fat pad, but upwards of 90% of their volume (Kanneganti and Dixit, 2012). The rest of the cell content is largely made up by immune cells, which can

play vital roles in AT function. Adipose macrophages for example, have been shown to cluster around dead or dying adipocytes and release proinflammatory cytokines (Weisberg et al., 2003, Cinti et al., 2005).

AT plays a very important role in appetite regulation by sending satiety and hunger cues, in the form of adipokines, to the central nervous system (CNS). Leptin, a major AT-secreted adipokine targets brain regions in the arcuate nucleus, which is one of the primary integrative centres for signals that regulate energy homeostasis (Woods and D'Alessio, 2008, Cone, 2005). Leptin is released from AT in proportion to its mass and acts hormonally by binding to its receptors within the CNS to trigger satiety and energy expenditure (EE) (Carbone et al., 2012, Cone, 2005). Null mutations in the mouse leptin gene ($Lep^{ob/ob}$) cause hyperphagia and extreme obesity (Zhang et al., 1994). Another adipokine, adiponectin, has a primarily metabolically protective role; it increases insulin sensitivity, induces the oxidation of fatty acids and suppresses glucose production in the liver (Goldstein et al., 2009). Adiponectin is highly expressed in SAT and increased VAT mass diminishes adiponectin levels in the blood (Turer et al., 2011). Consistent with a metabolically protective role, adiponectin-deficient mice exhibit worsened diet-induced insulin resistance (Maeda et al., 2002), whereas mice resulting from the

crossing of the $Lep^{ob/ob}$ mice with an adiponectin overexpressing line, show improved glucose homeostasis (Kim et al., 2007). Strikingly, this improved metabolic profile is not accompanied by weight-loss; these mice are the most corpulent murine models to date, indicating that adiponectin may act by promoting the preferential lipid storage in fat depots and limiting ectopic AT deposition in other tissues, therefore facilitating an improved systemic metabolic profile (Kim et al., 2007).

1.4.2 AT expansion and the ECM

AT has a remarkable capacity to expand and remodel, uniquely so among non-neoplastic tissues. AT expansion can happen in two ways. Firstly, through hyperplasia, when new adipocytes are recruited from available progenitors. Secondly, through hypertrophy, when existing adipocytes increase in size and content. In humans, upon sustained nutrient excess, different fat depots follow different paths to AT expansion. Upper body fat remodels primarily through hypertrophy, while lower body fat expands through hyperplasia (Tchoukalova et al., 2010). The same phenomenon is observed for VAT expansion in rodents (DiGirolamo et al., 1998).

Throughout this process, surrounding tissues need to accommodate this expansion. Namely, the newly formed or expanded adipocytes require nutrient and oxygen supply. AT depots are well vascularised for this purpose and new blood vessels are usually formed prior to expansion (Nishimura et al., 2007). Vascularisation potential varies between depots, with SAT having superior angiogenic potential in humans (Gealekman et al., 2011). However, adipocyte hypertrophy induces a hypoxic environment, which in turn triggers tissue remodelling and reorganisation of the extracellular matrix (ECM) (Pasarica et al., 2009, He et al., 2011). Therefore, hypoxia in AT expansion plays the role of a pro-inflammatory and pro-fibrotic signal.

More specifically, the ECM is responsible for supporting cell function and growth and in the case of adipocytes there is increased ECM deposition with hypertrophy; collagens, laminins, fibronectins and proteoglycans accumulate around the expanding adipocytes (Nakajima et al., 2002). This eventually leads to an imbalance of ECM deposition and degradation and causes the ECM to become dense, fibrotic and dysfunctional. This AT signature has also been linked with poor metabolic health (Spencer et al., 2010), but also with formerly obese individuals

who have undergone weight loss (Divoux et al., 2010). It is therefore unclear how and when fibrosis in AT becomes pathogenic.

1.5 Genetic underpinnings of obesity

Obesity is a partly heritable trait and as such, even though it is facilitated by today's obesogenic environments, genetic predisposition also plays an important role. Heritability estimates for obesity vary based on the measure used, but generally for BMI, twin-study based heritability ranges from 60 to 80% (Elks et al., 2012). While these estimates can sometimes be inflated (Manolio et al., 2009), due to demographic and study design confounders, they do indicate a significant genetic component to trait variation. Disentangling the genetics of obesity and particularly of detrimental AT deposition is of paramount importance for identifying high-risk individuals and delivering effective interventions.

The first insights into the genetics of obesity came from screening studies of severe human obesity. These revealed mutations in the leptin pathway (Montague et al., 1997, Vaisse et al., 1998). Such studies used methods like linkage analysis in affected pedigrees and were suited to detect rare variants of large, mostly monogenic, effects. The focus later shifted towards studying the common

variation of quantitative traits at the population level and in the case of obesity, mainly studying the variation in BMI. This initially happened via genome-wide linkage studies, which yielded no unequivocal associations (Saunders et al., 2007), or candidate gene studies, and later on via genome wide association studies (GWAS) that proved to be a much more powerful tool.

These earlier studies did not succeed at finding genetic signals, as they were ideally suited to detect rare and high-impact variants. Contrarily, tools designed to query common genetic variation in large population-based samples, like GWAS, were more useful for the detection of genetic signals associated with complex polygenic traits.

1.5.1 Early GWAS efforts

The strongest association with a potential obesity gene to date came in 2007, when the first major GWAS meta-analysis (GWAMA) was published (WTCCC, 2007). This highlighted single nucleotide polymorphisms (SNPs) within the gene *FTO*, that associated with incidence of T2D, while this association was later shown to be mediated through BMI (Frayling et al., 2007). Carriers for both copies of the effect allele were on average 3kg heavier than homozygous carriers of the protective allele.

The field of anthropometric GWAS was later dominated by the GIANT consortium. Early efforts consisted of GWAMAs of around 250,000 participants of European ancestry (Locke et al., 2015, Shungin et al., 2015) and collectively identified around 150 loci that associated with BMI and WHR or WHR adjusted for BMI (WHRadjBMI). As an anthropometric phenotype, WHRadjBMI has aptly been described as "unexpected fat distribution given the BMI" (Winkler et al., 2018).

These GWAMAs also showed that BMI loci were enriched for genes involved in the development of the brain and of the CNS (Locke et al., 2015), supporting a prominent role for appetite regulation in the genetic aetiology of obesity, similar to rare monogenic forms of the disease. Contrarily, WHR loci included distinct genes involved in peripheral metabolic pathways, such as adipocyte regulation and insulin resistance (Shungin et al., 2015).

1.5.2 Large biobanks and discovery power

In recent years, there has been an increase in the number of population-based biobanks in several countries. These include the Finnish cohort FINNGEN, the two American cohorts All of US and the Million Veteran Project, the Estonian Biobank and Biobank Japan, all of which have either already recruited or plan to recruit 500,000 to 1 million participants. Pertinent to this work, UKBiobank (UKB)

is a UK-based effort that includes genomic and phenotypic data for 500,000 middle-aged volunteers (Sudlow et al., 2015).

Recent GIANT GWAMAs included the data discussed above, alongside data from the UKB, bringing the collective cohort size to over 700,000 participants. This increase in size greatly enhanced the GWAS discovery power and yielded an astounding number of new associations. More specifically, the BMI GWAMA identified a total of 941 genome-wide significant (GWS) signals, with 751 of them being novel (Yengo et al., 2018). A BMI GWAS that only looked for rare coding mutations in the same individuals, identified rare variants in 13 genes, 8 of which had not been associated with human obesity phenotypes before (Turcot et al., 2018). The equivalent GWAMA for WHRadjBMI, identified 346 GWS loci, 300 of which were novel and a seventh of which appeared to also associate with body fat percentage (Pulit et al., 2018). The exonic GWAMA for WHRadjBMI identified 23 novel signals, 9 of which were rare (Justice et al., 2019). All four of these studies also supported the aforementioned dichotomy between BMI genes acting in the CNS and WHR genes acting in the metabolic periphery.

Overall, these GWAS identified over a thousand genomic loci that associate with anthropometric facets of adiposity, highlighting many known genes and

hundreds of novel ones. Translating these findings into biological mechanisms has lagged behind, as outlined below.

1.5.3 GWAS obesity genes & follow-up

In an effort to draw more direct biological insights, some of the GWAS above only surveyed how coding variation affects BMI and WHRadjBMI (Turcot et al., 2018, Justice et al., 2019). Theoretically, such exonic variants, and especially rare ones, should exert a larger effect on the studied traits, when compared to common genomic variation (Turcot et al., 2018). Indeed, a nonsense mutation in the repeatedly BMI-associated *MC4R* gene (p.Tyr35Ter), which is part of the leptin signalling pathway, was associated with an increase of 7kg in heterozygous carriers of the minor allele, which has a minor allele frequency (MAF) of 0.1% (Turcot et al., 2018).

However, this is not the case for most GWAS findings, as over 90% of associated SNPs are either intronic or intergenic (Zhang et al., 2014). This complicates any efforts to translate them into biological insights. Additionally, as sample sizes have increased with each new GWAMA, more variants of smaller effects have been discovered (Figure 1.3A). This is consistent with an omnigenic model of inheritance for complex traits, whereby GWAS top hits highlight genes

that directly affect traits. Then, with increasing GWAS sample size associated genes might be merely co-expressed with or regulate the trait-affecting genes, explaining their small phenotype effects (Boyle et al., 2017). At the same time, the trait variance explained cumulatively by these variants of small effect has been increasing incrementally and we can only account for 4% of the heritability in BMI (shown in Figure 1.3B and calculated in (Yengo et al., 2018, Locke et al., 2015)).

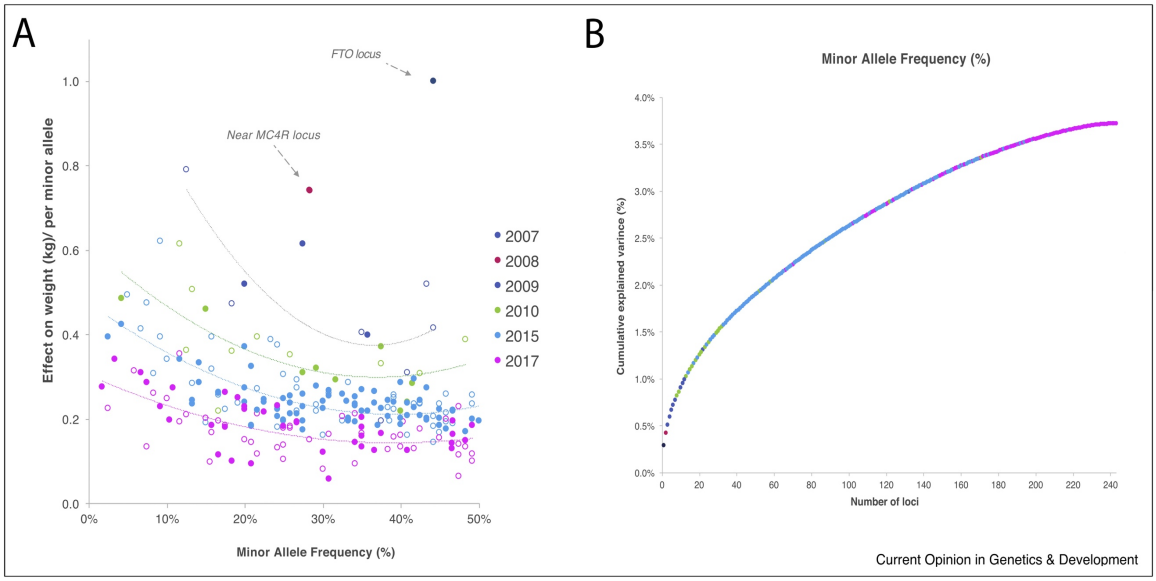


Figure 1.3 | Effect size and heritability of BMI-associated loci.

(A) Absolute effect sizes of the minor alleles, derived in kg and assuming a height of 170 cm, against the frequency of the alleles. Filled dots indicate a positive effect on BMI, hollow dots indicate a negative effect. (B) Cumulative amount of heritability explained with increasing number of loci. All coloured by year of publication/discovery. *Modified from (Loos, 2018).*

This all indicates two things. Firstly, that increasing cohort sizes and identifying more associated variants of minute effect, without understanding how they functionally affect the phenotype will not lead to actionable biological targets.

Disentangling the biological basis behind GWAS associations requires initially the identification of causal genes, which can often be challenging (further discussed in Chapter 3). Gene expression datasets can aid in this, by using them to identify how the associated variants affect expression in relevant tissues. *In vitro* models of these tissues or *in vivo* models of whole organisms can then be used introduce targeted alterations at the causal genes and characterise their effect (further expanded in Chapter 4).

Secondly, an important reason of the observed failure to account for the heritability of adiposity traits, is because these studies do not actually assay adiposity. While BMI and WHR are good proxies of body composition, using phenotypes that are closer to the adipose patterns themselves, could explain more of the observed trait variation. Many of the aforementioned cohorts have deployed whole body imaging, however due to the associated financial and labour costs, this has only happened for subsets of their participants (discussed in Chapter 2).

1.6 Aims & hypothesis

In the past decade, obesity associated loci have gone from tens (Herrera et al., 2011) to over a thousand (Yengo et al., 2018, Pulit et al., 2018). Even so, heritability estimates indicate that more genetic variation remains to be identified. More importantly, pinpointing the causal genes behind these associations and explaining how they exert their effect on body composition is greatly lagging behind. Largescale biobanks with detailed adiposity and anthropometric phenotyping, such as UKB, offer a fertile ground for mitigating this issue.

In this thesis, I exploited the availability of these resources to investigate *"The genetic underpinnings of adiposity: from GWAS discovery to functional characterisation".* To achieve this, I had the following three aims:

I. Identify genomic loci that associate with refined measures of human adiposity traits in UKB and replicate them (Chapter 2)
II. Investigate causal genes at these loci using gene expression datasets and chromatin dynamics (Chapter 3)
III. Describe how these genes affect body composition in an *in vivo* murine system (Chapter 4)

2

Identification of novel candidate loci for human adiposity

2.1 Introduction

2.1.1 Direct adiposity measures & GWAS

Imaging platforms such as MRIs, CTs and DXA scans offer accurate and direct measures of whole body and segmental adiposity (reviewed in Section 1.3.3). An effective way of studying whole-body adiposity measurements is scaling them to height-squared, creating indices similar to BMI (Heymsfield et al., 2007). This can be used to produce a lean mass index (LMI) and fat mass index (FMI). FMI is more sensitive and representative of true body composition than BMI and even raw body fat percentage (Kelly et al., 2009). Appendicular LMI is often used as an indicator of sarcopenia, i.e. the age-related degeneration of muscle mass and body strength (Iannuzzi-Sucich et al., 2002). Sarcopenic obesity, defined as excess body fat accompanied by a muscle deficit, is a particularly clinically important facet of metabolic disease and can be expressed as the fat to muscle mass ratio (FMR), over the whole body or segmentally (Ronco and De Stéfani, 2012).

Earlier GWAS studies of adiposity phenotypes using DXA scans identified novel adiposity associations (Liu et al., 2008), while also emphasising the sexually dimorphic nature of fat distribution (Fox et al., 2012). A more recent GWAMA of VAT, SAT and pericardial fat, was conducted in a collective cohort of just over 18 thousand individuals and identified seven novel loci (Chu et al., 2017). The first large GWAMA of DXA derived phenotypes (Neville et al., 2019), with a collective cohort size of 17,000 individuals, showed distinct genetic signatures for different fat depots. More specifically, Neville and colleagues showed that VAT-associated variants in their analyses exerted no effect over SAT and others had opposing effects on leg fat mass. A non-imaging study on the BIA phenotypes in UKB (Rask-Andersen et al., 2019) identified 29 novel loci, that associated with the proportions of AT in the general trunk, arm or leg areas. Studies such as these have highlighted the novel discovery potential of using direct adiposity measures in a GWAS framework. However, it is clear from their low sample sizes that large biobanks are still behind in implementing whole body imaging on their participants, when compared to anthropometry.

2.1.2 Phenotype imputation for increased power

Imaging-based adiposity phenotypes are expensive, time-consuming and require specialised medical equipment and staff to analyse them. Thus, population cohorts often prioritise anthropometry over imaging, as is the case with UKB, where anthropometric and BIA phenotypes are available for 500,000 participants, whereas DXA measures are only available for 5,000 participants. However crude they may be, anthropometry and BIA can still provide body composition insights as seen above. Thus, if it were possible to reconstruct a sufficiently accurate DXA phenotype from them, using statistical modelling approaches, could markedly increase discovery power in the context of a GWAS study and provide many novel candidate loci.

Such imputation efforts have followed varied approaches and have yielded some successes. For example, a study imputed BMI, triglycerides and systolic blood pressure, in a small Finnish cohort, using related phenotypes in a multivariate framework, but only managed to nominally replicate the GWAS findings obtained when using the original phenotypes for the same cohort (Hormozdiari et al., 2016). Other studies use participant and/or phenotype relatedness for phenotype imputation (Dahl et al., 2016, Joshi et al., 2017).

Together, the above studies demonstrated the potential for using imputed phenotypes in GWAS and beyond, when the trait of interest has only been measured for a subset of the cohort.

2.1.3 Aims

As seen in Section 1.2, measures of obesity and adiposity are highly correlated, often causally, with metabolic health and disease. Insights into these relationships are hindered by the lack of available direct measures of human adiposity. To address this, we aimed to:

I. Impute the DXA phenotypes onto all 500,000 UKB participants
II. Use them to perform GWAS in UKB
III. Replicate findings in independent cohorts with DXA measures

2.2 Experimental design & procedure

2.2.1 Human cohort data

2.2.1.1 The UKB cohort

UKB is a large population-based cohort with over 500,000 participants, aged 40 to 69, which were recruited in the UK over a period of 5 years, from 2006 to 2010. Its aim is to improve the prevention, diagnosis and treatment of serious and life-threatening illnesses affecting people of middle- and old-age. UKB

includes extensive phenotypic and genotypic data on its participants, including questionnaire data, physical measures, blood and urine sample assays, accelerometry, multimodal imaging, genome-wide genotyping and longitudinal follow-up (Sudlow et al., 2015). Phenotypes and genotypes were directly downloaded from the UKB website. All participants gave written informed consent and the study was approved by the North West Multicentre Research Ethics Committee.

Relevant to this study, participants underwent anthropometric examinations, of their height, weight, waist and hip circumference and BIA on a Tanita BC418MA body composition analyser. These measurements were available for the entirety of the cohort (493,088 participants). Participants' body composition was further assessed using the GE Lunar scanner, for a subset of 5,170 individuals. Scans of the whole body are analysed by the radiographer at acquisition to generate all numerical measures of bone mass and body composition. These measures are transferred directly from the instrument to UKB servers and require no post-processing.

UKB participants were genotyped under two Affymetrix arrays, which show a 96% SNP overlap and resulted in 820,967 genetic markers being genotyped.

SNPs were excluded on the basis of missingness and departure from Hardy-Weinberg equilibrium (HWE) (Bycroft et al., 2018). Imputation was performed by UKB, using SHAPEIT2 and IMPUTE2 on the UK10K, HRC and 1,000 Genomes Phase 3 reference panels (Sudlow et al., 2015). These imputed genotypes were downloaded for me in July 2017 and used for this work, under project #19655.

2.2.1.2 The DXA replication cohorts

2.2.1.2.1 ORCADES

ORCADES is a population-based isolate that includes 2215 individuals. Its aim is to characterise the genetic and epidemiological components that underlie quantitative traits and diseases in the Orkney Islands of Scotland. Individuals of all ages were recruited based on the basis of their Orcadian heritage and have at least two Orcadian grandparents, thus maintaining the homogeneous genetic background of the cohort. Data collection was carried out between 2005 and 2011 in Orkney by trained research nurses. The Orkney Research Ethics Committee and North of Scotland Local Research Ethics Committee granted the study ethical approval and all participants individuals gave written, informed consent prior to participating in any research, such as broad-ranging health and disease or

population research, including biobanking of samples or record linkage to hospital admissions or to other records (McQuillan et al., 2008).

A subset of 1,256 individuals of the ORCADES cohort underwent full body composition analysis on the Hologic fan beam DXA scanner (GE Healthcare). Trained radiology research nurses generated the scans and ensured the correct positioning of the participants' pelvis, arms and legs. The APEX2 software was used to generate individual measurements for bone, lean, and fat tissue of the head, arms, trunk, legs and total body. Subsequently, the APEX4 software was used to estimate the individuals' android, gynoid and visceral fat and lean mass content. These measures were derived by others and used in the context of this project.

DNA was extracted from whole blood and genotyped on three different SNP chips: Illumina Infinium Human Hap300v2, OMNI1 and OMNI Express. Each covers between ~300,000 and 1,000,000 variants across the genome, only ~160,00 of which overlap between the chips.

Genotypes were called via the Illumina BeadStudio and GenomeStudio software. Samples or SNPs with a call rate of under 97%, samples with a gender mismatch and SNPs deviating from HWE with a p-value of smaller than 1e-6 were excluded from all further analyses. Imputation was done separately for the Hap

and Omni chips, using the IMPUTE software (version 2.2.2) on the 1,000 Genomes European imputation panel and the exomes of 90 Orcadians. The two imputations were then merged, covering 37.5 million polymorphisms across the genome (Joshi et al., 2013).

2.2.1.2.2 EPIC-Norfolk

The European Prospective Investigation of Cancer (EPIC) is a large multi-centre prospective cohort study that focuses on the connection between diet, lifestyle factors and cancer. The Norfolk cohort includes 30,000 individuals, aged 40 to 79, living in Norwich and the surrounding towns and rural areas. The participants were recruited between 1993 and 1997 and have been contributing information about their lifestyle and health through questionnaires and health checks for over two decades. All participants gave signed informed consent and The Norwich District Health Authority Ethics Committee approved the study (Day et al., 1999).

A subset of the cohort (~4100 participants) was DXA scanned at the fourth health check, 20 years after initial recruitment, at the EPIC-Norfolk Unit at the Norwich Community Hospital. The Lunar Prodigy advanced fan beam scanner (GE Healthcare) and the enCORE software version 14.10.022 (GE Healthcare) were

used. Participants were scanned by trained operators, using standard imaging and positioning protocols. The enCORE software was used to demarcate the regional boundaries. All the images were manually processed by one trained researcher, who corrected demarcations according to a standardized procedure (Lotta et al., 2017).

The EPIC samples genotyped using the Affymetrix GeneChip Human Mapping 500K Array Set and genotypes were called by using the BRLMM algorithm. SNPs were subject to the following quality control (QC) criteria: Hardy–Weinberg p-value > 10e-6, MAF>5% and SNP call rate > 95% (Soranzo et al., 2009, Ong et al., 2009). In the context of this project, we requested 27 SNPs (Table 2.1).

2.2.1.2.3 Fenland

The Fenland Study is an ongoing population-based cohort study, that includes 12,434 adults aged 29-65 years in Cambridgeshire, UK. The first phase of the study investigates the interaction between environmental and genetic factors determining obesity, T2D, and related metabolic disorders. Volunteers were recruited from general practice registers between 2005 and 2015. A follow up study (Phase 2) was launched in 2014 with the objective of studying the relationship between change in objectively quantified behaviours and body

composition and metabolic risk. The Fenland study was approved by the Cambridge Local Research Ethics Committee and all participants gave written informed consent (De Lucia Rolfe et al., 2010).

Body composition by DXA was assessed, using the Lunar Prodigy Advanced fan beam scanner (GE Healthcare) with a constant pixel size of 1.2 mm. Estimates of total body fat mass and total abdominal fat (g) were derived with Prodigy enCORE software (version 10.51.006; GE Healthcare). The DXA abdominal fat region (g) was defined by quadrilateral boxes with the base of the box touching the pelvis and the lateral boundaries extending to the edge of the abdominal soft tissue (Mytton et al., 2018, De Lucia Rolfe et al., 2010).

Participants were genotyped on the Affymetrix SNP5.0 chip and genotypes were called using BRLMM. Variant inclusion criteria were MAF > 1%, call rate > 90% and HWE p-value >10e-6. Imputation was conducted via the IMPUTE2 software and all SNPs with an info-score >0.4 were kept (Winkler et al., 2015). In the context of this project, we were provided with 27 SNPs (Table 2.1).

Table 2.1 | Loci of interest to be replicated in the DXA cohort meta-analysis (MA).

The information submitted for replication to the external DXA cohorts. Phenotype calculations and explanations can be found in Table 2.3. Genomic locations given in GRCh37. Gene annotations made simply by proximity (+/- 500kb) to the lead associated variant using dbSNP build 150 and a1 denotes the effect allele.

rsid	*snpid*	*a1*	*Gene*	*Phenotype*
rs4820325	22_38599978	A	*MAFF/PLA2G6*	Leg FMR
rs117068593	14_93118229	T	*RIN3*	Leg FMR
rs1047891	2_211540507	A	*CPS1*	Android LMI
rs150985439	2_172417398	C	*CYBRD1*	Android LMI
rs62218301	21_36741262	G	*LOC100506403*	Leg FMR
rs13072731	3_38533335	C	*ACVR2B/EXOG*	Android LMI
rs17684944	17_43187141	C	*NMT1/MIR6784*	Leg FMR
rs2736208	8_116826244	C	*TRPS1*	Gynoid fat mass
rs11079849	17_47090785	T	*IGF2BP1*	SFI
rs17325374	2_230822858	G	*FBXO36*	Android FMI
rs3749748	5_127350549	T	*LINC01184*	Android LMI
rs12359330	10_72414845	T	*ADAMTS14*	Leg FMR
rs12451882	17_42301922	T	*UBTF*	Total lean %
rs4527444	4_30842780	G	*PCDH7*	Android LMI
rs1504797	8_89434405	C	*MMP16*	Android LMI
rs2881198	7_46634506	C	*LOC730338*	Android LMI
rs10761129	9_94486321	T	*ROR2*	Android LMI
rs2242449	17_7095507	T	*DLG4*	Total lean %
rs4832298	2_86764004	T	*CHMP3*	Leg FMI
rs7893571	10_16750129	T	*RSU1*	Android fat mass
rs11525873	7_138817193	C	*ZC3HAV1/TTC26*	Android LMI
rs6557791	8_21949663	C	*FAM160B2*	Leg FMR
rs12274114	11_68255577	C	*LRP5/PPP6R3*	Android LMI
rs1441264	13_79580919	A	*LINC00331*	SFI
rs2504236	13_28613298	T	*FLT3*	Android LMI
rs4660586	1_42407229	T	*HIVEP3*	Gynoid fat mass
rs59893724	5_80830788	G	*SSBP2*	SAT

Table 2.2 | Population cohorts used for replication.

In descending sample size, the 4 DXA replication cohorts. Individuals included in the N below, make up the final cohorts after all exclusions took place. Cohorts marked as external were given the 27 SNPs in Table 2.1, above, to replicate. Analyses in internal cohorts were conducted by myself.

Cohort	*Type*	*N*
Fenland	Replication – External	8034
UKB$_{DXA}$	Replication – Internal	4366
EPIC-Norfolk	Replication – External	4134
ORCADES	Replication – Internal	1253
Total Replication	–	**17787**

2.2.2 Phenotype Prioritisation & Imputation

In March 2018, derived DXA phenotypes were available for a subset of 5,170 individuals within the UKB study. These phenotypes included fat mass, lean mass, bone mass and total mass, for the following seven general body areas: android, gynoid, arm, leg, trunk, VAT and total body. Phenotypes were downloaded and used for the purposes of this work, alongside anthropometric (height, weight, waist and hip circumference), BIA (fat, fat-free and total mass for the arms, legs, trunk and total body, basal metabolic rate and total water weight), demographic (sex, age, assessment centre, Townsend deprivation index, educational attainment, Northing and Easting, genetic ethnicity) and disease phenotypes

(incidence of CVD, T2D, high cholesterol, hand grip strength and walking pace as measures of physical fitness). Metabolic syndrome was also calculated, using the criteria in (NCEP, 2001), but excluding blood biomarker levels as they were not available in UKB at the time this work took place. For the physical measures, where these phenotypes had been recorded at multiple instances, the measurements were averaged. Related and non-white-British individuals were excluded from all analyses.

The remaining UKB participants were separated into two sub-cohorts; the UKB_{DXA} cohort (approximately 4,400 participants), used to prioritise and impute the DXA phenotypes (iDXA) onto the UKB_{iDXA} cohort (approximately 390,000 participants), which was used for subsequent genetic discovery analyses. The UKB_{DXA} cohort was also part of the DXA replication MA cohort. Apart from the derived DXA phenotypes, we also created composite phenotypes and indices to approximate adipose distribution patterns (calculated as outlined in Table 2.3). Phenotypic correlations between all pre-derived and calculated DXA measures and the aforementioned disease phenotypes were used to determine which body composition patterns were the best predictors of metabolic health or disease.

Prioritised DXA phenotypes were sex-separated and imputed, using linear models that incorporated all of the available anthropometric and BAI phenotypes, as mentioned above and seen below,

$$DXA = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_n x_n + \varepsilon$$

where, β is the effect of each x phenotype on the measured DXA,

β_0 is the intercept of the linear model,

ε is the normally-distributed residual error and

the proxies (x) used were Age, Height, Waist circumference, Hip circumference, WHR, Weight, BMI, BIA total fat %, BIA total fat mass, BIA total fat-free mass, BIA total water mass, BIA basal metabolic rate, BIA trunk fat %, BIA trunk fat mass, BIA trunk fat-free mass, BIA total trunk mass, BIA leg fat %, BIA leg fat mass, BIA leg fat-free mass, BIA leg total mass, BIA arm fat %, BIA arm fat mass, BIA arm fat-free mass and BIA arm total mass.

Effect estimates for each of these model components were then used to estimate the iDXA phenotypes in the individuals which had anthropometric and BIA phenotypes measured but lacked the DXA ones. To assess imputation efficacy, the iDXA phenotypes were correlated to the original DXA ones and also to the disease phenotypes used for prioritisation.

2.2.3 GWAS methodology

All of the sex-separated iDXA phenotypes were natural log-transformed to achieve a normal distribution, prior to GWAS. Following that, iDXA values lying

further than six standard deviations on either side of the population mean were removed. They were corrected for age, assessment centre, geographical coordinates, Townsend deprivation index, educational attainment, the first 20 principle components (PCs), genotyping array and batch, as is standard practise and as seen below. QQ plots were visually inspected to confirm that the covariates used to adjust for population stratification were sufficient to control type-I error rate. Non-white-British and related individuals were excluded from all genetic analyses.

iDXA phenotype ~ Age + Assessment centre + Northing + Easting + Townsend deprivation index + Educational attainment + Genotyping batch + Genotyping array + PC1 + … + PC20

The resulting residuals were inverse rank transformed and values lying four standard deviations at either side of the population mean were removed. Residuals from men and women were merged and used for GWAS. For each SNP, the iDXA phenotypes were regressed on the three possible genotypes, in REGSCAN (Haller et al., 2015), assuming additive genetic effects. SNPs with a MAF <0.001 and an imputation quality score <0.4 were excluded from analyses.

2.2.4 Multiple testing correction

To correct for multiple testing, the R package PhenoSpD (Zheng et al., 2017) was used to ascertain the effective number of tests conducted, given the phenotypic similarity of the iDXA phenotypes, which was four. The genome wide significance threshold was then adjusted accordingly and set at 1.25×10^{-8}.

All SNP associations below this threshold were compiled across all iDXA phenotypes, sorted into separate signals based on their location (in 1Mb windows) and assessed for novelty, checking for previous associations with any adiposity or anthropometry phenotypes in the GWAS Catalog, under the key word "obesity" (accessed April 2018) (Buniello et al., 2018). SNPs within published loci and SNPs which were associated with BMI or WHR (p-value$<10^{-12}$) were excluded.

2.2.5 Replication approach

Independent novel loci were followed up with a replication power calculation assuming the effect sizes in the iDXA GWAS would be equal to those in the DXA analyses, for a sample size of 18,000 DXA participants. This includes the collective DXA individuals of EPIC-Norfolk, Fenland, UKB_{DXA} and ORCADES (Table 2.2). Loci were prioritised for replication based on this replication cohort

size and on whether they would pass a 5% false discovery rate (FDR), under the assumption that the iDXA and DXA effect sizes would be the same.

Genotype and DXA phenotype information were submitted to analysts from EPIC-Norfolk and Fenland as outlined in Table 2.1. I conducted the UKB_{DXA} and ORCADES analyses in the same manner. The DXA phenotypes were transformed in the same way the iDXA were (Section 2.2.3) and were used to test for association with allelic dosage for the selected 27 iDXA SNPs. Replication was considered successful when the DXA FDR-corrected two-sided p-value was <0.05 and the direction of effect was consistent with the one observed in iDXA.

2.3 Results

2.3.1 iDXA phenotypes were predictive of adiposity and metabolic health

DXA phenotypes were measured in a subset of UKB (5,107 participants) and regressed against all available anthropometric and BIA phenotypes. These models were used to predict the DXA phenotypes in over 400,000 UKB participants. The Pearson correlation coefficients observed between the DXA and iDXA phenotypes, for all individuals within this UKB_{DXA} sub-cohort were overall R=0.81 and generally

indicated accurate predictions of the underlying phenotypes (Figure 2.1C, Table 2.3 and Figure S1).

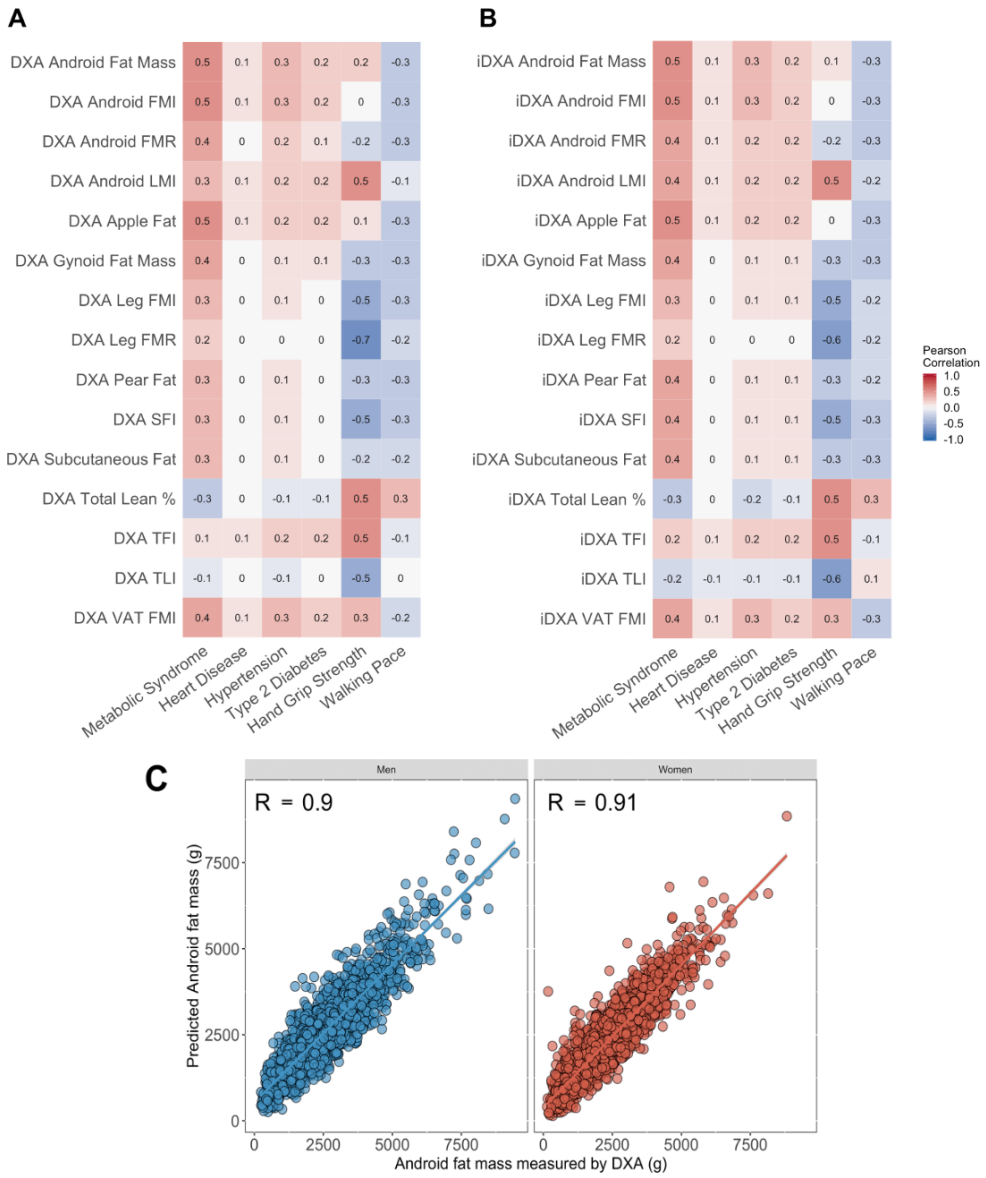


Figure 2.1 | iDXA phenotypes were accurate predictions of DXA and of metabolic health.

Phenotypic correlations between obesity disease sequelae and DXA (A) and iDXA (B) phenotypes. Sex separated correlations with disease can be found in Figure S2. (C) Example iDXA sex-separated prediction models as a function of DXA-measured Android fat mass. Predictions for all other iDXA phenotypes can be found in Figure S1 and R values can be found in Table 2.3.

The iDXA phenotypes remained as predictive as the DXA phenotypes of metabolic health (Figure 2.1A&B). Phenotypes representing android adipose distribution patterns correlated positively with incidence heart disease and T2D, while gynoid patterns of AT distribution did less so. However, some phenotypes indicative of limb fat mass correlated strongly with reductions in physical fitness, as measured by hand grip strength and walking pace.

Table 2.3 | Correlations between measured DXA and iDXA phenotypes.

Pearson coefficients (R) calculated for the 5,000 participants with both DXA and iDXA, separately for the two sexes, ranged from 51 to 91%, but were 81% on average. Height measures used in cm and mass in grams.

Phenotype	*Calculation*	*Sex*	*R*
Android Fat Mass	Pre-calculated	women	0.91
		men	0.90
Android FMI	Android fat mass divided by height squared	women	0.91
		men	0.89
Android FMR	Android fat mass divided by android lean mass	women	0.87
		men	0.83
Android LMI	Android lean mass divided by height squared	women	0.74
		men	0.71
Apple Fat	Total fat mass subtracting leg fat mass	women	0.92
		men	0.91
Gynoid Fat Mass	Pre-calculated	women	0.90
		men	0.89
Leg FMI	Leg fat mass divided by height squared	women	0.86
		men	0.83
Leg FMR	Leg fat mass divided by leg lean mass	women	0.77

Pear Fat	Total fat mass subtracting trunk and arm fat mass	men	0.73
		women	0.86
SFI	Segmental fat index; the sum of arm and leg fat divided by height squared	men	0.85
		women	0.88
SAT	Android fat mass subtracting visceral adipose tissue	men	0.85
		women	0.88
Total Lean %	Total lean mass divided by total mass	men	0.69
		women	0.87
TFI	Trunk fat index; trunk fat mass divided by the sum of arm and leg fat mass	men	0.84
		women	0.71
TLI	Trunk lean index; trunk lean mass divided by the sum of arm and leg lean mass	men	0.58
		women	0.53
VAT FMI	VAT mass divided by height squared	men	0.51
		women	0.84
		men	0.82

2.3.2 Discovery approach yields hundreds of novel loci

The predicted iDXA phenotypes were used to perform GWAS in the extended UKB cohort, excluding the 4,400 individuals that were used in the iDXA estimations. Cumulatively, the GWAS yielded just under 5,000 GWS signals (4,764 SNPs), at the adjusted threshold of $p<1.25x10^{-8}$. Among these, many were well known and had been previously associated with BMI and other traits, such as *FTO, MC4R, RSPO3, TMEM18* and others.

Thus, the list of loci was checked against all known "obesity" associations on the GWAS catalog and for discoverability for BMI, WHR in the entire UKB cohort.

After exclusions, these mapped to 1,251 non-overlapping loci (Figure 2.2, Table S2). The remaining loci were screened for any residual known associations or locus overlap and visually checked. 242 loci were considered sufficiently novel and could be taken forward to replication.

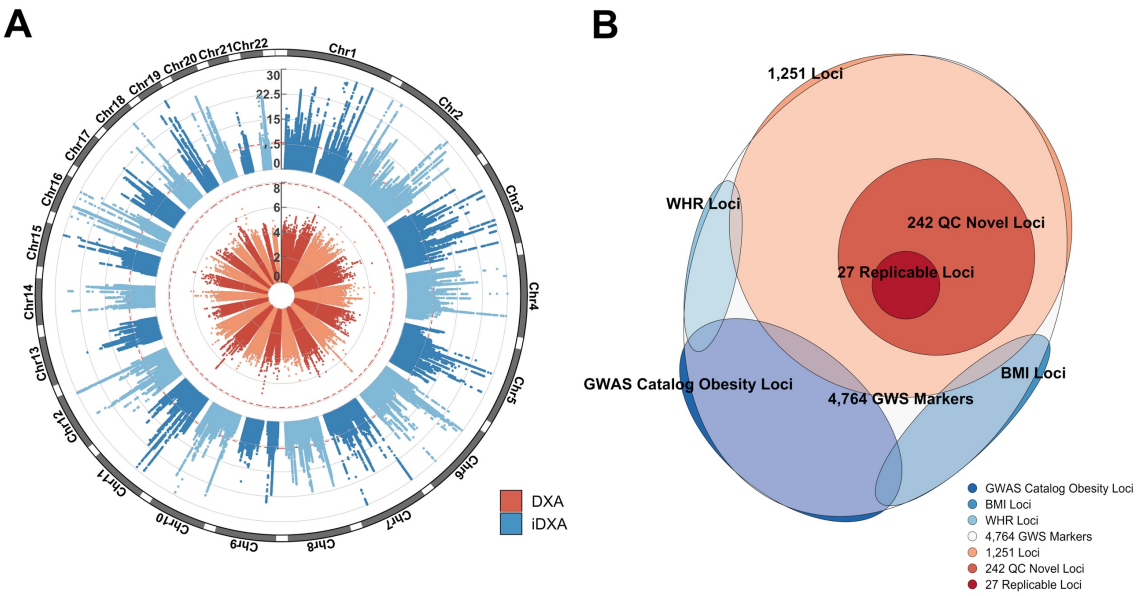


Figure 2.2 | iDXA approach leads to the discovery of hundreds of novel adiposity loci.

(A) GWAS loci associated with collated DXA and iDXA phenotypes from Table 2.3 in the UKB_{DXA} and UKB_{iDXA} cohorts, respectively. iDXA p-values capped at 10^{-30} for visibility. Red line indicates the GWS threshold, set at $1.25x10^{-8}$. Manhattan and QQ plots for each GWAS can be found in Figure S3. GWAS model coefficients can be found in Table S1. (B) Visualisation of the filtering and exclusion of identified loci based mainly on known associations with other phenotypes. Cumulative number of GWS SNPs (4,764) across all iDXA GWAS were screened for known associations on the GWAS Catalog and with BMI and WHR. The remaining SNPs were mapped to 1,251 independent loci, which were further filtered for known associations. This left 242 novel loci, 27 of which were replicable for a replication sample size of 18,000 DXA-measured participants. SNP information for all these loci can be found in Table S2.

2.3.3 Six out of 27 loci replicate successfully in DXA

The replication meta-analysis (MA) cohort consisted of 18,000 participants from 4 cohorts. Of the 242 novel loci identified in the iDXA GWAS, 27 would be replicable (p<5%) with N=18,000, if the iDXA effect sizes were true. Thus, the 27 strongest iDXA associated variants were submitted for replication in the corresponding DXA phenotypes. At an FDR of 10%, 6 loci replicated successfully (P<5%), while the majority of the variants (19/27) were directionally consistent between DXA and iDXA (Figure 2.3).

The first replicated SNP, rs4820325 (Figure 2.3B), was located proximally to *PLA2G6* and within *MAFF,* in a gene-rich region of chromosome 22 and was the strongest association to the iDXA Leg FMR phenotype. Per copy of the effect allele (rs4820325-A), leg fat mass was reduced by 19 grams and conversely leg lean mass was increased by 40 grams. Other variants in linkage disequilibrium (LD) with the index SNP have been suggestively associated, but not at GWS, with body fat percentage (Lu et al., 2016) and also many other metabolic traits (Spracklen et al., 2017, Teslovich et al., 2010, Willer et al., 2013). The second variant, rs1047891 (Figure 2.3C), is a missense SNP in the *CPS1* gene and was associated with a 7 gram increase in lean mass in the android area per copy of the effect allele (rs1047891-

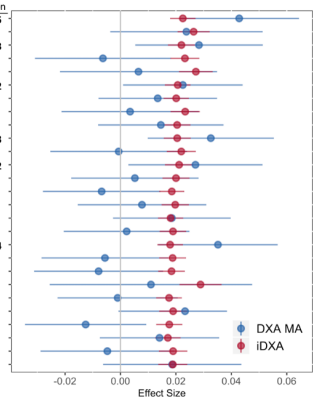
A). This locus was previously associated with renal and metabolic phenotypes (Pattaro et al., 2016, Van Meurs et al., 2013). The third variant, rs13072731 (Figure 2.3D), is located in *ACVR2B* and showed a 6 gram increase in android lean mass per copy of the effect allele (rs13072731-C). The locus was previously associated with various electrocardiogram traits (Newton-Cheh et al., 2009). The fourth, rs17325374 (Figure 2.3E), was associated with an increase of 25 grams in android fat mass per copy of the effect allele (rs17325374-G) and was located between genes *FBXO36* and *TRIP12*. The fifth variant, rs12359330 (Figure 2.3F), is located in gene *ADAMTS14* and exhibited an increase in leg fat mass by 18 grams or a reduction in leg lean mass by 37 grams, per copy of the effect allele (rs12359330-T). Previous GWAS linked this locus to circulating white blood cell counts (Astle et al., 2016). The final variant is located in gene *DLG4* (Figure 2.3G), rs2242449-T was associated with an 85 gram reduction of total lean mass, per copy of the T allele. This locus was previously associated to liver enzyme and blood metabolite levels (Kettunen et al., 2012, Chambers et al., 2011).

Figure 2.3 | Six out of 27 prioritised loci replicated in the DXA-MA at 10% FDR.

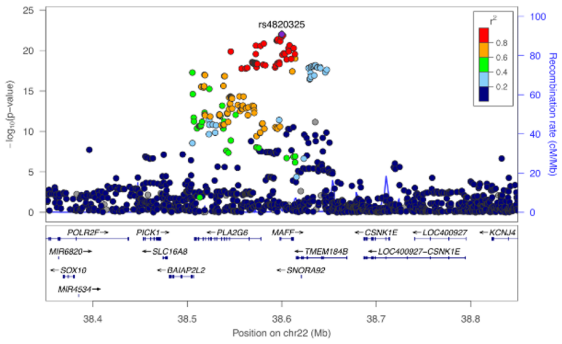
(A) SNP information for the 27 prioritised loci in Discovery-iDXA and Replication-DXA. Where, a1 is the effect allele and freq1 is the frequency of the effect allele in Discovery. Gene column annotated simply by proximity to the lead associated variant. Effect sizes in the iDXA analyses are presented as absolute values and also given for all loci in Table S2 and for the lead variants at the PLA2G6 and ADAMTS14 loci across all GWS iDXA traits in Table S4. (B-G) Locus zoom plots for the six replicated variants and surrounding loci (+/-500kb). Genomic locations and annotations given in GRCh37/hg19 and dbSNP build 150.

A

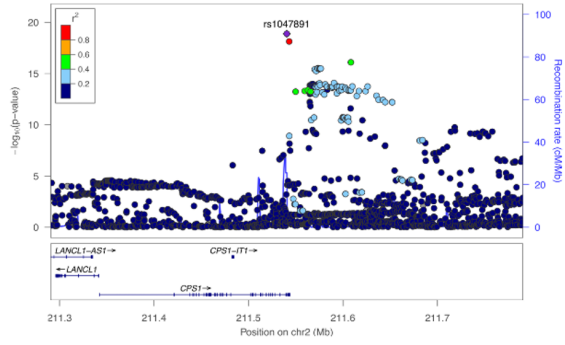
rsid	Phenotype	Gene	snpid	a1	freq1	Discovery	Replication
rs4820325	**Leg FMR**	**MAFF/PLA2G6**	**22_38599978**	**A**	**0.58**	**9.61E-23**	**3.54E-05**
rs117068593	Leg FMR	RIN3	14_93118229	T	0.19	7.25E-20	0.04
rs1047891	**Android LMI**	**CPS1**	**2_211540507**	**A**	**0.32**	**1.31E-19**	**6.89E-03**
rs150985439	Android LMI	CYBRD1	2_172417398	C	0.25	3.50E-19	0.70
rs62218301	Leg FMR	LOC100506403	21_36741262	G	0.17	3.69E-19	0.32
rs13072731	**Android LMI**	**ACVR2B/EXOG**	**3_38533335**	**C**	**0.61**	**5.15E-19**	**1.85E-02**
rs17684944	Leg FMR	NMT1/MIR6784	17_43187141	C	0.55	9.80E-19	0.10
rs2736208	Gynoid fat mass	TRPS1	8_116826244	C	0.76	9.83E-19	0.39
rs11079849	SFI	IGF2BP1	17_47090785	T	0.33	2.33E-17	0.10
rs17325374	**Android FMI**	**FBXO36**	**2_230822858**	**G**	**0.32**	**3.58E-17**	**2.08E-03**
rs3749748	Android LMI	LINC01184	5_127350549	T	0.25	6.00E-17	0.52
rs12359330	**Leg FMR**	**ADAMTS14**	**10_72414845**	**T**	**0.27**	**1.32E-16**	**1.27E-02**
rs12451882	Total lean %	UBTF	17_42301922	T	0.31	2.06E-16	0.33
rs4527444	Android LMI	PCDH7	4_30842780	G	0.54	3.20E-16	0.74
rs1504797	Android LMI	MMP16	8_89434405	C	0.30	9.00E-16	0.25
rs2881198	Android LMI	LOC730338	7_46634506	C	0.54	1.82E-15	0.04
rs10761129	Android LMI	ROR2	9_94486321	T	0.67	3.34E-15	0.42
rs2242449	**Total lean %**	**DLG4**	**17_7095507**	**T**	**0.43**	**6.94E-15**	**5.48E-04**
rs4832298	Leg FMI	CHMP3	2_86764004	T	0.69	1.10E-14	0.69
rs7893571	Android fat mass	RSU1	10_16750129	T	0.66	1.51E-14	0.75
rs11525873	Android LMI	ZC3HAV1/TTC26	7_138817193	C	0.10	2.88E-14	0.27
rs6557791	Leg FMR	FAM160B2	8_21949663	C	0.60	3.44E-14	0.54
rs12274114	Android LMI	LRP5/PPP6R3	11_68255577	C	0.28	5.18E-14	0.03
rs1441264	SFI	LINC00331	13_79580919	A	0.59	6.02E-14	0.87
rs2504236	Android LMI	FLT3	13_28613298	T	0.59	1.11E-13	0.10
rs4660586	Gynoid fat mass	HIVEP3	1_42407229	T	0.74	1.82E-13	0.65
rs59893724	SAT	SSBP2	5_80830788	G	0.25	6.00E-13	0.07



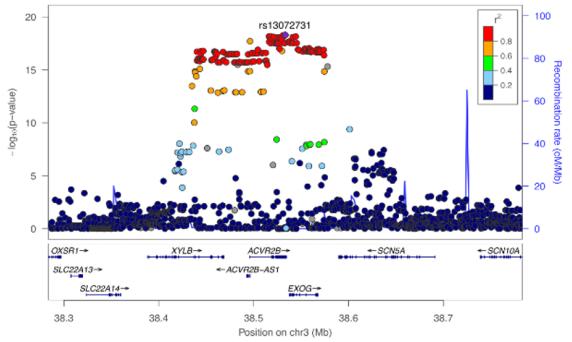
B



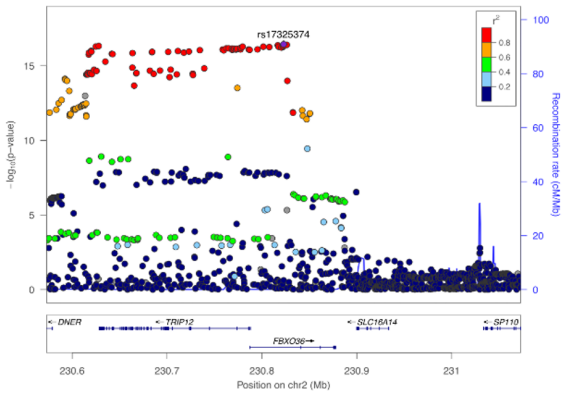
C



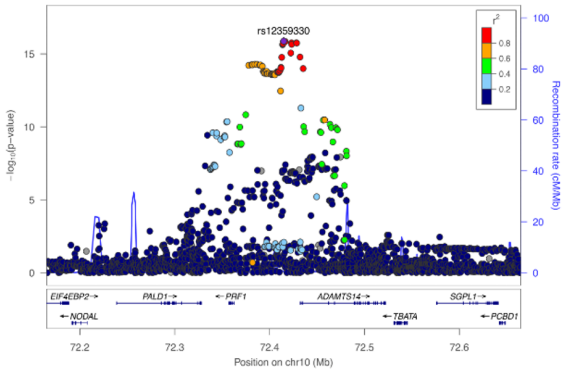
D


rs13072731
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
OXSR1→
XYLB→
ACVR2B→
←SCN5A
←SCN10A
SLC22A13→
←ACVR2B-AS1
SLC22A14→
EXOG→
38.3
38.4
38.5
38.6
38.7
Position on chr3 (Mb)

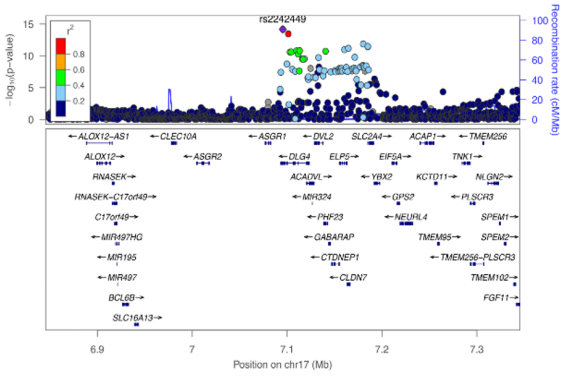
E


rs17325374
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
←DNER
←TRIP12
←SLC16A14
←SP110
FBXO36→
230.6
230.7
230.8
230.9
231
Position on chr2 (Mb)

F


rs12359330
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
EIF4EBP2→
PALD1→
←PRF1
ADAMTS14→
SGPL1→
←NODAL
←TBATA
←PCBD1
72.2
72.3
72.4
72.5
72.6
Position on chr10 (Mb)

G


rs2242449
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
←ALOX12-AS1
←CLEC10A
←ASGR1
←DVL2
SLC2A4→
ACAP1→
←TMEM256
ALOX12→
←ASGR2
←DLG4
ELP5→
EIF5A→
TNK1→
RNASEK→
ACADVL→
←YBX2
KCTD11→
NLGN2→
RNASEK-C17orf49→
←MIR324
←GPS2
←PLSCR3
C17orf49→
←PHF23
←NEURL4
SPEM1→
←MIR497HG
←GABARAP
TMEM95→
SPEM2→
←MIR195
←CTDNEP1
←TMEM256-PLSCR3
←MIR497
←CLDN7
TMEM102→
BCL6B→
FGF11→
SLC16A13→
6.9
7
7.1
7.2
7.3
Position on chr17 (Mb)

2.4 Discussion

In this chapter, anthropometric and BIA proxies were used to impute direct DXA measures of adiposity and body composition in the UKB_{DXA} cohort. These predictions provided reasonably accurate estimates of the underlying DXA phenotypes and of consequent metabolic disease risk, and were further used in the UKB_{iDXA} cohort to perform GWAS. This yielded over a thousand GWS loci (1,251) and 242 novel ones, while the equivalent GWAS in UKB_{DXA} yielded no GWS associations. A MA was formed by four DXA cohorts to replicate these novel loci and out of 27 tested signals, 6 were successfully replicated. Two of these genes were chosen for experimental follow-up, for the reasons outlined below.

2.4.1 iDXA promise & limitations

Accurate whole-body and segmental adiposity measures are needed in order to fully delineate the genetic drivers of body composition and its contributions to disease. To this end, several population cohorts have invested in whole-body scans of their participants, be it MRI, CTs or DXA scans. GWAS in these cohorts have started to provide some insights into the genetics of human body composition (Fox et al., 2012, Chu et al., 2017, Neville et al., 2019). However, these are lagging far behind the GWAS signals for traits like BMI and WHR (Pulit et al.,

2018, Yengo et al., 2018). This is largely due to the fact that even the largest DXA MA to date (Neville et al., 2019) was less than one 35th the size of the largest BMI MA (Yengo et al., 2018), i.e. <20,000 versus 700,000 participants, due to the expense associated with scanning this many people. Importantly, the loci emerging from these analyses show generally little overlap, emphasising the limitation of BMI for informing on the genetics of adiposity. Thus, viable alternatives are needed to derive detailed adiposity measures.

To this end, we developed a simple linear prediction method that uses body composition proxies to calculate imputed phenotypes for DXA, iDXA. This method utilised the breadth and depth of cohorts such as UKB; basing its predictions on both anthropometric and BIA phenotypes and quantifying their contributions to the DXA phenotypes in the small UKB_{DXA} sub-cohort, before imputing them onto the larger UKB_{iDXA} cohort. Doing this, we showed that the iDXA phenotypes correlated well with the DXA ones (R=0.81, Figure S1) and that they remained predictive of the incidence of metabolic disease (Figure 1.1). This indicated that, overall, they were accurate representations of adiposity.

GWAS on the iDXA phenotypes in UKB_{iDXA} gave rise to just under five thousand (4,764) GWS signals and 242 novel loci (Figure 2.2). In an effort to not

only replicate these associations, but also verify the prediction method, we sought replication in the MA of 4 separate DXA cohorts. Out of the 27 SNPs submitted for replication, 6 replicated at an FDR-corrected p-value<5%. These are further discussed in Section 2.4.2, below.

The DXA imputation approach increased our discovery power greatly, however it is important to consider the following issues. Firstly, while the imputation method worked reasonably well for most DXA phenotypes, some predictions were poor (i.e. R=50-70%, see Table 2.3). This seemed to be an issue for DXA indices that were based around a lean mass phenotype (i.e. Android LMI, Leg FMR and TLI). This was to be anticipated, as none of the proxy phenotypes used in the models offer direct measures of lean mass. When using BIA, lean mass is often underestimated in lean subjects and overestimated in obese subjects (Fuller, 1993, Fuller et al., 1994), which could help explain the prediction issue. More importantly, BIA individual measures are always calculated using equations suited to the population or study sample. Applying the same equations to a young versus aging population increases bias, especially for lean mass, in the older participants (Scafoglieri et al., 2017).

Secondly, with any similar approach, there is the concern of whether the genetic signals were associations to the underlying DXA phenotypes or merely to some of the components of the models we used. While this would be impossible to preclude without having a much larger DXA cohort to test, arguably even if that were the case, the resulting associations would still provide novel knowledge towards human body composition. Finally and interestingly, the 6 replicated SNPs all had larger effect sizes in DXA than they did in the iDXA analysis. Meaning, that the iDXA GWAS underestimated the effect of the variants on the phenotypic variation (Figure 2.3A). While the opposite phenomenon (i.e. inflation of the effect sizes) would have been more problematic, this remains something to consider as a potential artefact of the method.

2.4.2 *Pla2g6* and *Adamts14* loci were prioritised for follow-up

Six of the 27 prioritised iDXA loci were successfully replicated in the measured DXA phenotypes (Figure 2.3). These loci associated with various adiposity phenotypes and had quite small effects on them, as covered in Section 2.3.3. When prioritising loci for experimental follow-up, the candidates were judged predominantly in terms of their novelty and the biological plausibility of the putative causative genes at the loci (expanded in Table S3).

Additionally, certain aspects of some of the loci made them unsuitable. For example, the sixth locus (near genes *ACADVL*|*DLG4*) is very gene-rich, which could complicate any fine-mapping attempts due to the sheer number of genes at that locus. *ACADVL,* the gene within which the top SNP lies, has also been extensively studied in a metabolic context (Cox et al., 2001, Exil et al., 2006). Therefore, it was unlikely that this locus would be particularly tractable or novel. The third locus, near genes *ACVR2B*|*EXOG*, had similar issues; even though less gene rich, the gene where the index SNP was located, *ACVR2B*, was a well-known regulator of body composition (Zaragosi et al., 2010, Guo et al., 2009), questioning the likely novelty of any experimental findings. For the second locus, the top SNP (rs1047891) was a missense variant (4217C>A, turning residue 1406 of the protein to Asparginine from Threonine) in gene *CPS1*, increasing the likelihood of *CPS1* being the causal gene at the locus. However, mice with homozygous null mutations in *CPS1* exhibit postnatal lethality due to hyperammonaemia (Schofield et al., 1999), which would complicate *in vivo* characterisation for this gene with heterozygous crossings. Similarly for the fourth locus, the top SNP was located between *TRIP12* and *FBXO36* and knock-out (KO) models for both genes exhibit severe developmentally impaired phenotypes (Kajiro et al., 2011, Dickinson et al., 2016).

The first locus was also quite gene rich (Figure S4), but it had been previously nominally associated with total body fat percentage (Lu et al., 2016), making it a more established GWAS candidate. The gene just upstream of the top SNP, *PLA2G6*, encodes a protein which catalyses the release of fatty acids from phospholipids. GWAS have previously associated PLA2G6 SNPs which are in moderate or high LD with rs4820325 with circulating triglyceride levels (Spracklen et al., 2017, Teslovich et al., 2010, Willer et al., 2013), making it a gene of interest for metabolic dysfunction. Mice with null mutations in *Pla2g6* exhibit impaired male fertility and dystrophy of axons and synapses in the brain (Shinzawa et al., 2008). Inactivation of the gene attenuates liver steatosis and improves glucose tolerance in $Lep^{ob/ob}$ mice (Deng et al., 2016). *Pla2g6* deficiency does not cause dyslipidaemia in mice exposed to chronic high fat diet (HFD) feeding, supporting a metabolically favourable role for the KO of *Pla2g6 (Zhang et al., 2013)*. In the iDXA GWAS there is reduced leg fat mass by 19 grams or increases leg lean mass by 40 grams, per copy of the effect allele (rs4820325-A). Notably, other significant SNPs at this locus associate with increased expression of *PLA2G6,* in terms of mRNA levels *(Lonsdale et al., 2013)*. This indicates that increased expression of *PLA2G6* was associated with increased lean mass or reduced fat mass, which is

contradictory to the mouse phenotypes. Thus, arguably, the GWAS signal could be driven by one of the many other genes at this locus, making causal gene identification pivotal to understand this association.

Finally, the index SNP at the fifth locus fell within the gene *ADAMTS14* (Figure S4), a metalloproteinase and aminoprocollagen peptidase (further explored in Section 4.1.2). Mutations in this gene were previously associated with osteoarthritis and GWAS have linked unrelated SNPs at this locus to lymphocyte and reticulocyte counts (Astle et al., 2016, Rodriguez-Lopez et al., 2009). *Adamts14*$^{-/-}$ mice reportedly had no significant phenotype (Dupont et al., 2018). However, metalloproteinases are known regulators of development and body composition (further explored in Section 4.1.1), making *Adamts14* a plausible candidate within the context of altered adiposity.

Given, the *PLA2G6* and *ADAMTS14* loci were selected to be followed-up in different experimental systems. To try and understand the mechanism behind a well-established metabolic GWAS signal for the former and potentially uncover a novel driver of body composition for the latter.

2.5 Conclusion

In summary, we developed a linear regression method that successfully imputed DXA measures in UKB, using anthropometry and BIA. After showing that the iDXA phenotypes were accurate representations of the underlying DXA ones, we used them to perform GWAS. This caused a marked increase in discovery power and led to the identification of hundreds of novel loci associated with body composition. Of the 27 strongest associations, 6 loci replicated successfully in independent DXA cohorts. Out of these, the *PLA2G6* and *ADAMTS14* locus were be followed-up, initially to identify the causal genes that drive the associated phenotypes (Chapter 3) and further to understand how they affect body composition (Chapter 4).

3

GWAS signals & causal gene assignment

3.1 Introduction

3.1.1 GWAS variants and gene regulation

As reviewed in Section 1.5, GWAS of anthropometry and obesity traits have identified over a thousand genomic associations (Pulit et al., 2018, Yengo et al., 2018). Over 90% of these signals are either intronic or intergenic, making their functional significance difficult to delineate (Dong et al., 2018). Often, with GWAS loci, the gene closest to the associated SNPs is assigned as causal. However, this is often an oversimplification.

Using the archetypal BMI-associated locus, *FTO* (Frayling et al., 2007), as a paradigm, the GWAS signal lies within an intron of the *FTO* gene. Thus, early efforts focused on *FTO* as the causal gene. A mouse model with a large deletion including the *FTO* gene is embryonically lethal, primarily due to abnormalities in brain development (Anselme et al., 2007), while mice with more targeted deletions of exons 2 and 3, exhibit less severe growth retardation and decreased fat and lean mass, albeit with high post-natal lethality, somewhat backing the role of *FTO* as

the gene driving the observed association (Fischer et al., 2009). However, humans with equivalent inactivating mutations exhibited no weight-related phenotype (Boissel et al., 2009), indicating that perhaps the severe developmental mouse phenotype is unrelated to the GWAS results.

Using direct DNA chromatin interactions at the *FTO* locus, Smemo and colleagues showed that the association interval forms long-range interactions with the promoter of gene *IRX3*, in mouse, human and zebrafish tissues (Smemo et al., 2014). Further to this, using a battery of computational and functional approaches, Claussnitzer and colleagues showed that certain associated variants within the first intron of *FTO* actually affect the binding of distal enhancers of *IRX3* and *IRX5*, activating them specifically in adipocyte precursors (Claussnitzer et al., 2015). They also showed this activation was dependent on the presence of the risk-conferring variants.

On the other hand, the *FTO* locus SNPs with the largest effects on BMI co-localise with transcriptional regulators of gene *RPGRIP1L* (Stratigopoulos et al., 2008) and that these variants associate with changes in the expression of *RPGRIP1L*, specifically in neurons (Stratigopoulos et al., 2014), further complicating the interpretation of the *FTO* variant effect on body weight.

This exemplifies the need for detailed fine-mapping of GWAS loci. Genomic signals are merely the first step towards truly understanding the causal relationships between genome variation and quantitative traits.

3.1.2 Methods for gene assignment in GWAS

There are several methods for assigning causal genes within a GWAS candidate locus. For the past few years, many methods that look at the physical chromatin interactions between gene elements and non-coding variation have been developed. Most of these operate under the assumption that this variation affects the binding efficacy between *cis*- or *trans*-acting enhancers and their target gene promoter, which is facilitated by the looping of chromatin (Benabdallah and Bickmore, 2015). Such methods, generally fall under the chromatin conformation capture (3C) umbrella and rely on snapshotting chromatin interactions within an area, tissue and stage of interest. DNA from cells is crosslinked chemically and then digested into small fragments, which represent genomic regions that were interacting within that population of cells. These fragments are further sequenced to identify their location within the assayed region.

Alternatively, expression quantitative trait locus (eQTL) analyses can also help with deciphering the way in which the identified genomic variants exert their

effects on the expression genes within a locus. The most expansive eQTL database is the GTEx database (Lonsdale et al., 2013), which is unique in the fact that it includes human gene expression data across most tissues, obtained from cadavers. Other efforts have focused on gene expression data in the blood, as these are much more easily assayable from live study participants, and could be arguably representative of whole body expression patterns (Westra et al., 2013, Lloyd-Jones et al., 2017, Võsa et al., 2018).

Such datasets can be used to check whether top associated SNPs are eQTLs for neighbouring genes. Several tools have also been developed that use GWAS summary statistics and eQTL data, to directly test the molecular mechanisms via which genomic variation can affect quantitative phenotypes (Gamazon et al., 2015). Many such tools use MR to test for pleiotropy between variation in gene expression and the trait of interest, in an effort to make causal inferences as to the gene which drives the phenotype. Such methods include summary-data-based MR (SMR) and transcriptome-wide summary statistics-based MR (TWMR) (Zhu et al., 2016, Porcu et al., 2019). Both have been used increasingly on GWAS results and are starting to highlight new causal genes.

3.1.3 Aims

In Chapter 2, we performed GWAS on iDXA phenotypes in UKB, to identify genomic regions that are associated with body composition. Having identified and replicated many novel loci, we decided to focus on two: the *PLA2G6* and the *ADAMTS14* locus. However, it remains unknown how the variants at these loci drive the associated phenotypes. Here, we investigated the effect of the associated intervals on gene regulation using publicly available datasets and direct chromatin evidence.

3.2 Experimental design & procedure

3.2.1 *In silico* methods

3.2.1.1 eQTL analysis

The effect of the top SNPs on gene expression was interrogated using the publicly available gene expression database GTEx. GTEx includes cadaver samples from 48 non-diseased tissue types from nearly 1000 individuals (Lonsdale et al., 2013).

3.2.1.2 SMR & TWMR

SMR and Heterogeneity in Independent Instruments (HEIDI) was used on the iDXA summary statistics to implicate causal gene candidates (Zhu et al., 2016),

where eQTL analyses did not suffice. SMR-HEIDI was conducted on gene expression data from GTEx (Lonsdale et al., 2013) to causally link gene expression and body composition. SMR tests whether an exposure of interest (i.e. mRNA levels) is associated with an exposure of interest (i.e. a quantitative trait), as mediated by an instrument variable constructed using genetic variant information. A significant association between trait and gene expression levels could be driven by a single causal variant (pleiotropy) or multiple linked variants that may be affecting the chosen exposure and outcome independently (linkage). The HEIDI test differentiates between the pleiotropy and linkage, under the null hypothesis that there is a single causal variant driving the association, based on whether the association pattern across multiple cis-SNPs within the associated region is homogeneous. Thus, SMR results were filtered at a <5% FDR threshold and >5% HEIDI-test p-value.

Further to that, TWMR was also used (Porcu et al., 2019), using specifically blood gene expression data from the blood of the >31,000 participants of the eQTLGen Consortium (Võsa et al., 2018). TWMR operates under similar MR assumptions, however it uses multiple independent instruments, each constructed by eQTLs of different genes, in an effort to minimise pleiotropy confounding.

3.2.2 Chromosome Conformation Capture Carbon Copy (5C)

3.2.2.1 Single cell suspension preparation

Wild type (WT) 3-month-old C57BL/6 mice were sacrificed humanely. The brain and testes were harvested and neuronal cortex and seminiferous tubule cells were microdissected and kept on ice. The tissues were then dissociated, by being placed in 1ml of TrypLE (ThermoFisher) and incubated at 37°C for 7 minutes, being vortexed every minute. After the seven minutes, they were triturated 10 times using a P1000. TrypLE was deactivated by addition of 1ml of phosphate buffered saline (PBS) supplemented with 2% foetal bovine serum (FBS) and then a further dilution up to 5ml. The dissociated cells were then passed through a 40nm filter to eliminate larger clumps and cell number was quantified using a TC20 Automated Cell Counter (BioRad). Samples were then separated into 10 million cell aliquots and gently pelleted at 200g for 10 minutes.

3.2.2.2 3C Library preparation

The pellets were then fixed in PBS containing 37% formaldehyde and incubated at room temperature (RT) for 10 minutes. After the incubation, glycine was added to stop the cross-linking for 5 minutes at RT followed by 15 minutes on

ice. Fixed cells were pelleted again at 4°C for 10min at 400g. Supernatants were aspirated and the pellets were snap frozen on dry ice and kept at -80°C.

To digest the chromatin, frozen pellets were thawed on ice and resuspended in 200µl of lysis buffer (10mM Tris-HCl pH 8.0, 10mM NaCL, 0.2% NP40 in water) with 20µl of protease inhibitor and incubated for 15 minutes on ice. Cells were disrupted on ice by douncing 40 times with Pestle B. Lysates were centrifuged for 5 minutes at 5000rpm at RT. The supernatant was discarded and pellets were resuspended in 100µl of restriction buffer and centrifuged for 5 minutes at 2000g. This was done twice and the pellets were finally resuspended in 50µl of the restriction buffer. 227µl of restriction buffer and 38µl of 1% SDS were added to each aliquot and incubated for 10 minutes at 65°C, after which 44µl of 10% Triton were added. Finally, 400U of EcoRI-HF (NEB) restriction enzyme were added per sample and incubated overnight at 37°C.

Samples were incubated at 65°C for 30 minutes to in activate the enzyme. To ligate the DNA, 97µl of ligation mix (1X T4 DNA ligase buffer, 10% Triton X-100, 0.1mg/ml BSA, 1mM ATP, 3000U T4 ligase) was added to each reaction and incubated for 4 hours at 16°C. To purify the ligation reactions, 15µl of 10mg/ml Proteinase K was added and incubated overnight at 65°C and then again for two

hours the next day. Equal volumes of phenol were added and the reactions were vortexed for 2 minutes and then centrifuged for 10 minutes at 1500g until the phases separated. The aqueous phase was transferred to a fresh tube and this was repeated twice with equal amounts of phenol/chloroform.

Then, 0.1 volumes of 3M sodium acetate (pH 5.2) and 2.5 volumes of cold absolute ethanol were added. Reactions were incubated at -80°C for at least 1hr or overnight and then centrifuged at maximum rpm for 25 minutes at 4°C. The supernatants were aspirated and the pellets were washed with 70% ETOH. The pellets were then resuspended in 50µl of 1X TE (pH 8.0) and incubated at 37°C for 15 minutes with 1µl of 10mg/ml RNAse A. Purified 3C libraries were then stored at -20°C or verified by gel electrophoresis.

3.2.2.3 5C primer and library design and preparation

5µg of 3C library were added to 10µg of salmon sperm DNA on ice and diluted with deionised water to a final volume of 10 µl. 5C primers were diluted tenfold on ice and fixed to a give a final concentration of 2nM. Mixed primers were added on ice to library samples with 1µl of annealing buffer (10X NEBuffer 4, New England Biolabs Inc.). Samples were then denatured for 5 minutes at 95°C and

annealed overnight at 48°C. Taq DNA ligase was added and the samples were incubated at 48°C for one hour and at 65°C for 10 minutes to stop the ligation.

2-5µl of ligated samples were used to amplify the 5C library the A-key and P1-key primers for 32 cycles at 60°C. Amplified 5C libraries were purified, using the AMPure beads (Beckman), quantified using a bioanalyser (Agilent) and submitted to sequencing with an Ion Torrent PGMTM Sequencer (LifeTechnologies).

3.2.2.4 5C data analysis

Biological replicates were sent for sequencing for both the testis and brain cells. Due to Covid-19 work restrictions, sequencing could not be completed on the second brain replicate and thus analysis was conducted on only one sample for the brain and two for the testis. The testis replicates were averaged and primers with very high trans-interactions were removed. The sequenced data was processed using the my5C website (my5c.umassmed.edu) (Lajoie et al., 2009) and were binned and smoothed over 15 kb windows. Genes and ENCODE genomic regulatory elements within the region were procured and visualised via the UCSC browser (http://genome.ucsc.edu/) (Rosenbloom et al., 2012).

3.3 Results

To establish causal genes at the *PLA2G6* and *ADAMTS14* loci, the strongest associated SNP at each locus was initially checked against eQTLs for genes in each region (1Mb windows). Further to this, SMR-HEIDI was used to look for causal effects of changes in gene expression on the iDXA body composition phenotypes, that match the pattern of SNP association. Both of these analyses were conducted using the GTEx tissue-wide expression data (V7, (Lonsdale et al., 2013)).

The *PLA2G6* locus was further analysed using TWMR on blood gene expression data from eQTLGen (Porcu et al., 2019) and further under a 5C framework. These chromatin assays aimed to observe interactions between variants within the GWAS peak in an approximately 650kb genomic domain and were performed in mouse-derived cells, due to the unavailability of human ones.

3.3.1 *ADAMTS14* is a strong causal gene candidate at the fifth locus

To attempt to identify a causal gene at this locus, the index SNP at this locus was checked against the eQTL and gene expression data for the GTEx portal (Lonsdale et al., 2013). This showed that the lead SNP, rs12359330, was indeed an eQTL for the expression of *ADAMTS14* and that the T allele associated with an increase in the expression of the gene (Figure 3.1C).

This was further corroborated by the SMR analysis, which matched the pattern of association with the pattern of gene expression across 48 tissue types and showed that *ADAMTS14* was a causal gene at this locus (Figure 3.1B).

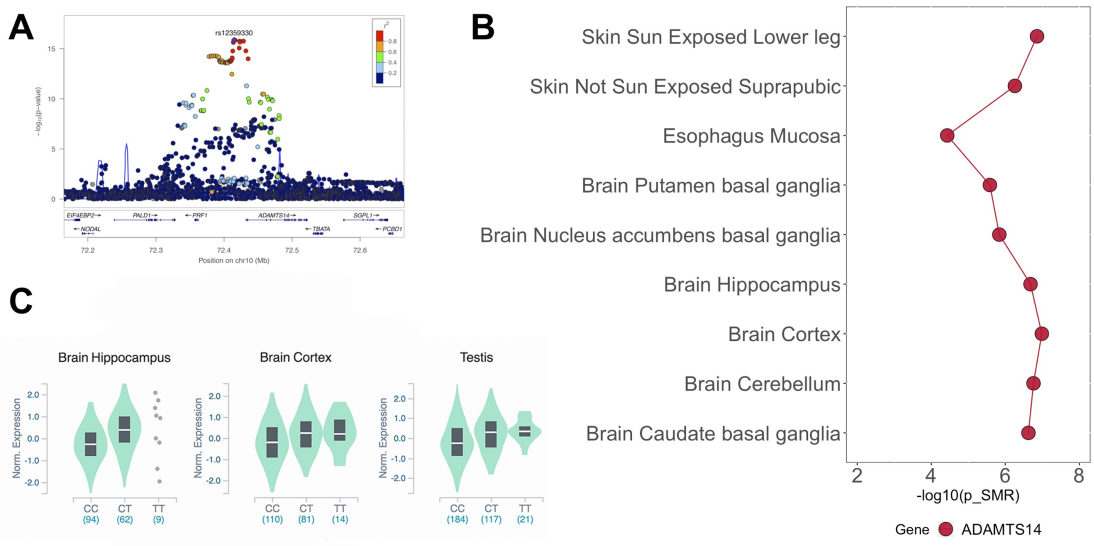


Figure 3.1 | eQTL and SMR data highlight *ADAMTS14* as the causative gene at this locus.

(A) Locus zoom plot, showing the associated variants at this locus, with rs12359330, located just upstream of the gene. (B) SMR p-values for all GTEx tissues at which the observed phenotype was caused by the observed pattern of association. (C) Direct eQTL data for the index SNP and a sample of tissues from the GTEx portal (Lonsdale et al., 2013).

3.3.2 Multiple genes might be causal at the *PLA2G6* locus

The eQTL analysis at the *PLA2G6* locus showed that the index SNP, rs4820325, was as eQTL for many genes at this locus (Figure S5 and Table S5). SMR using in the 48 GTEx tissues showed the same, changes in expression of seven genes were causally affecting the associated phenotype (Figure 3.2B). Using only

blood eQTLs, showed that changes in expression of 18 genes could mediate the observed change in the associated phenotype (Figure 3.2C).

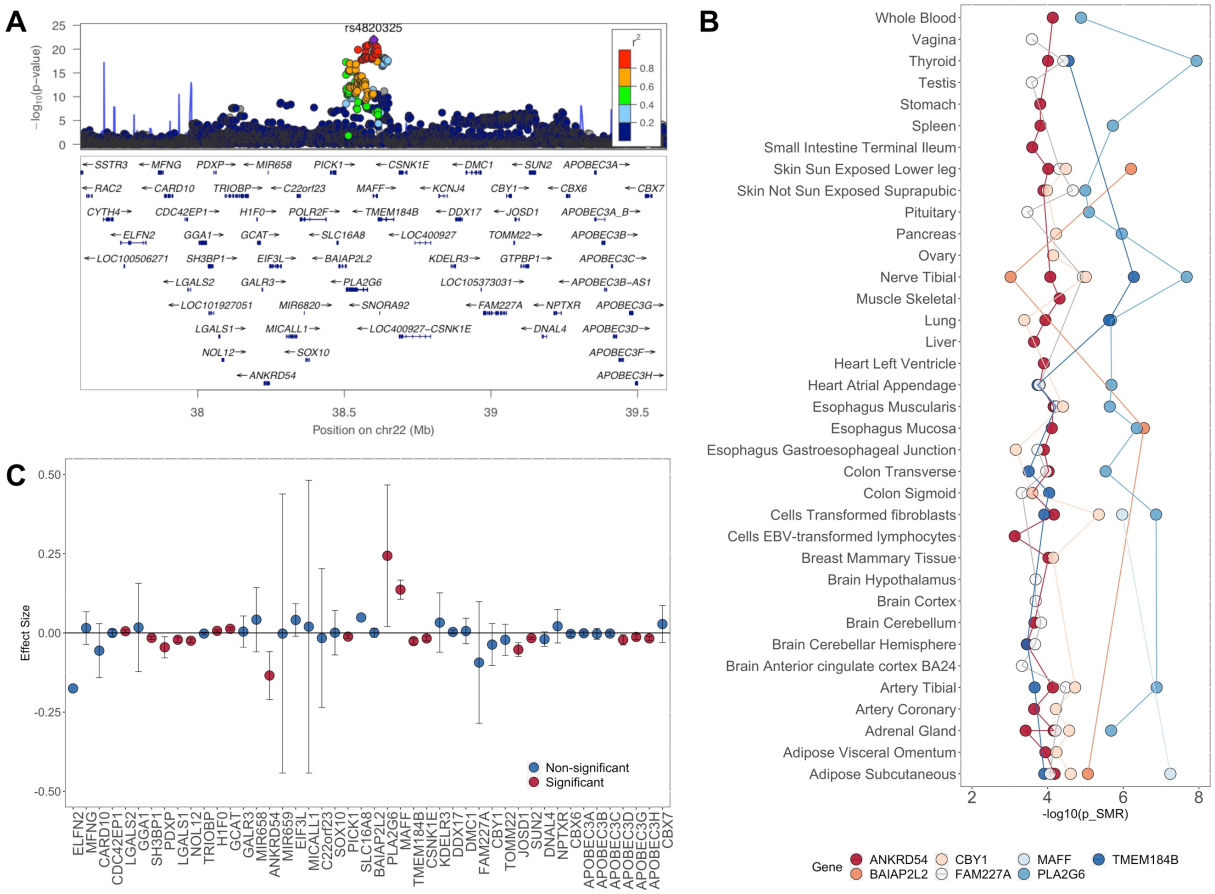


Figure 3.2 | SMR and TWMR causally implicate many genes at the *Pla2g6* locus.

(A) Locus zoom plot, showing the associated variants at this locus, with rs4820325 located nearby *PLA2G6* and *MAFF*. (B) SMR results for all GTEx tissues at which the observed phenotype was caused by the observed pattern of association. (C) TWMR on blood eQTLs showing significant genes at the wider locus (+/- 500kb from the index SNP), in chromosomal order.

3.3.3 5C data shows several physical interactions at the *PLA2G6* locus

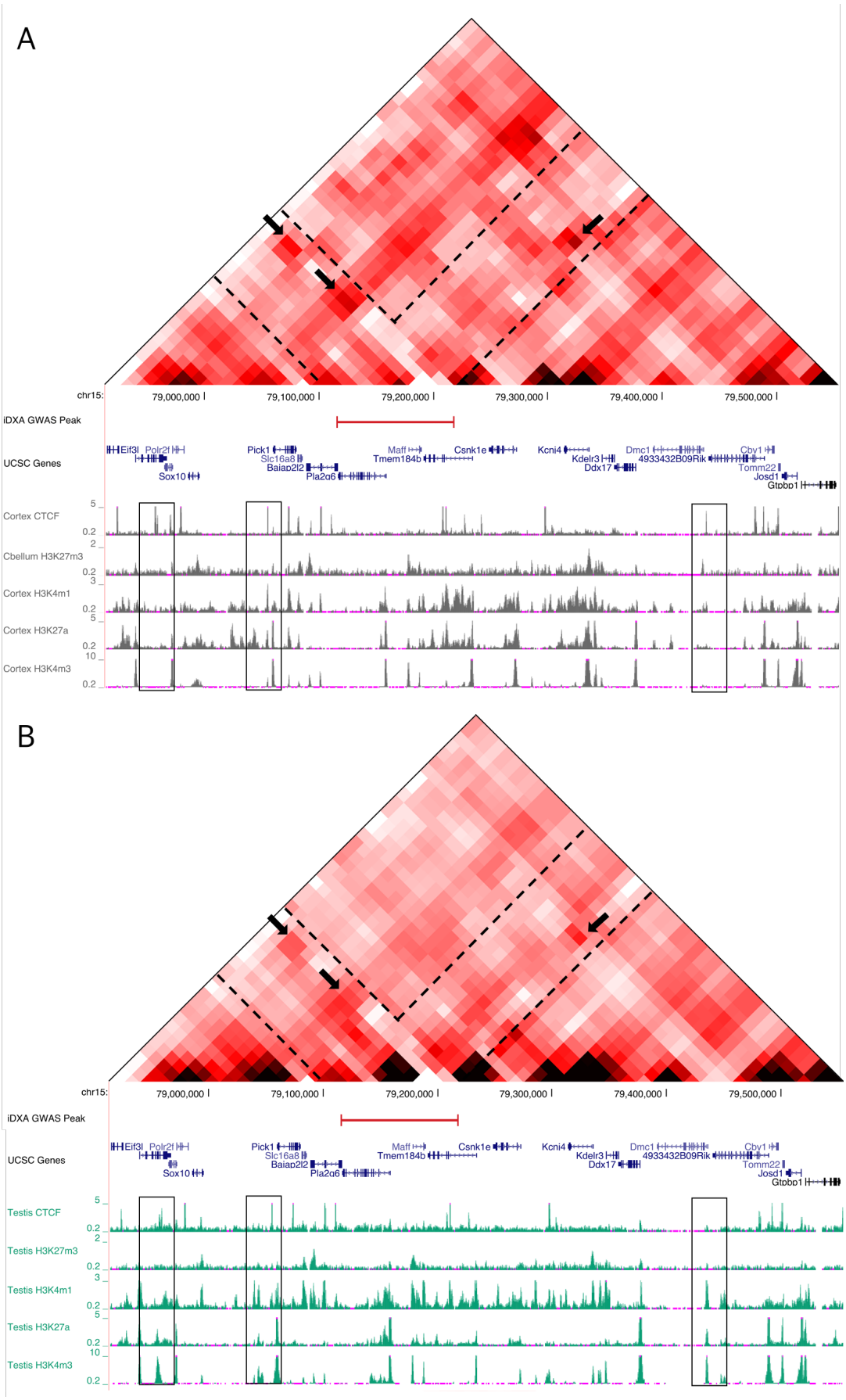
Given the number of candidate causal genes highlighted in the eQTL analyses, chromatic conformation analyses were used to try and narrow down the

number of gene candidates in the locus. Targeted 5C assays were carried out in a 650kb wide area, encompassing the 100kb of GWAS-highlighted SNPS (Figure 3.3). This was done in mouse-derived brain cortex and testis cells and for each cell type interaction frequencies were normalised based on a downstream desert, within chromosome 15.

The 5C data showed enriched chromatin interactions, in both cell types but at varying degrees, between the associated interval and three other genomic regions. Firstly, with the upstream region of *Pick1*, where there were strong CTCF peaks in both brain and testis cells. Secondly, with the upstream region of *Dmc1* which also coincides with CTCF sites and other histone marks in both tissues. Finally, with *Sox10* in a region of poised chromatin for the testis. All these interactions are further explored below.

Figure 3.3 | Physical chromatin interactions at the Pla2g6 locus.

Chromatin interactions within the 650kb window on chromosome 15 in mouse-derived brain cortex cells (A) and testis cells (B). Interaction frequencies were normalized to an unchanging gene desert region on chromosome 15 and binned and smoothed over 15 kb windows. GWAS peak area indicated in red and diagonal interactions with it are shown within the black dashed lines. Black arrows indicate interaction enrichment. UCSC genes and ENCODE tracks given for CTCF sites (repression), H3K27me3 (repression), H3K4me1 (poised enhancer), H3K27a (active enhancer) and H3K4me3 (active promoter) marks from 8-week-old mouse tissues. Normalised heatmaps, without binning and smoothing can be found in Figure S6 in the Appendix. (Overleaf)


A
chr15:
79,000,000
79,100,000
79,200,000
79,300,000
79,400,000
79,500,000
iDXA GWAS Peak
UCSC Genes
Eif3l
Polr2f
Sox10
Pick1
Slc16a8
Baiap2l2
Pla2g6
Maff
Tmem184b
Csnk1e
Kcni4
Kdelr3
Ddx17
Dmc1
4933432B09Rik
Cby1
Tomm22
Josd1
Gtpbp1
Cortex CTCF
Cbellum H3K27m3
Cortex H3K4m1
Cortex H3K27a
Cortex H3K4m3
B
chr15:
79,000,000
79,100,000
79,200,000
79,300,000
79,400,000
79,500,000
iDXA GWAS Peak
UCSC Genes
Testis CTCF
Testis H3K27m3
Testis H3K4m1
Testis H3K27a
Testis H3K4m3

3.4 Discussion

In this chapter, computational and experimental techniques were used to identify causal genes at the prioritised loci from Chapter 2. For the *ADAMTS14* locus, we were able to show that SNPs within the association interval were eQTLs for the expression of the *ADAMTS14* gene. This indicates that changes in the expression of *ADAMTS14* could causally drive the phenotype. The *PLA2G6* locus, contained several eQTLs for many different genes, making the identification of a causal candidate difficult. Long-range chromatin interactions within the interval of association showed that it co-localised with regulatory elements of three different genes, *Pick1*, *Dmc1* and *Sox10*.

3.4.1 *Adamts14*; future work & *in vivo* characterisation

The eQTL analyses highlighted *ADAMTS14* as a causal gene candidate at this locus. Namely, they showed homozygous carriers for the rs12359330-T allele exhibit increased expression of *ADAMTS14* and have 36 more grams of fat or 74 less grams of lean mass on their legs. Thus, this would be a good candidate to study in the context of body composition.

The mouse is a commonly used model organism in studies of metabolic disease (Yazdi et al., 2015), for many reasons. Firstly, they are easy to breed and

genetically manipulate. The mouse and human genome is also highly conserved, for *Adamts14*, the mouse protein is 82% similar to the human one (Table S3 via (Boratyn et al., 2012)), indicating the mouse is a suitable model for genetic perturbations, that may delineate conserved biological function. In terms of body composition, mice are the only animal model that has multiple distinct fat depots that share such similarities with the human ones (Chusyd et al., 2016) (Figure 3.4).

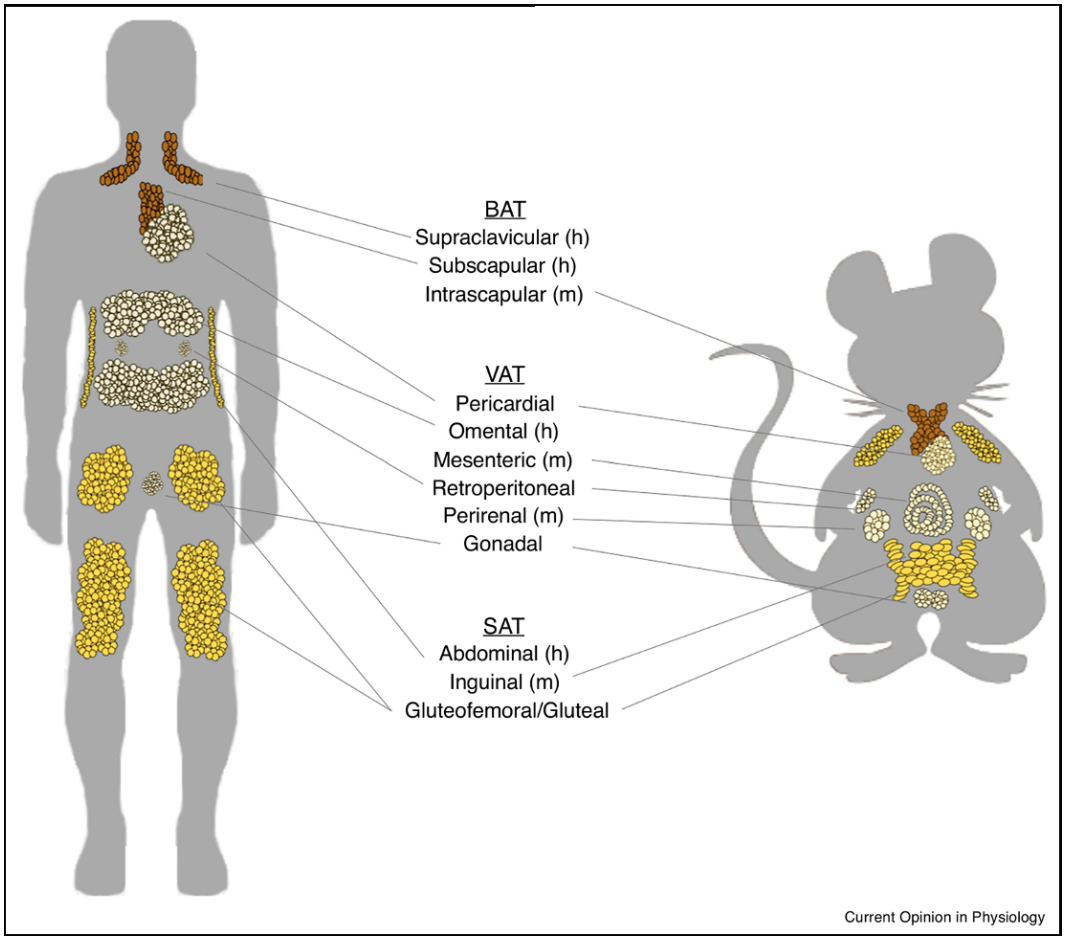


Figure 3.4 | Comparable anatomical distribution of fat depots in mice and humans.

Both species are multi-depot animals and have BAT depots around the scapular area. VAT depots are localised around and between the intra-abdominal organs. SAT is located primarily on the lower halves of the body. *Taken from (Kentistou et al., 2019).*

The human VAT depot can be equated to the mesenteric, retroperitoneal, gonadal and perirenal pads, while SAT equates to the anterior and posterior pads in mice (Steel et al., 2005). When studying the metabolic implications of the expansion of different types of AT, like in the current study, using a multi-depot model is critical.

Finally, it is worth noting that, even though in this case, the route from genomic variation to gene expression to phenotype seemed straight-forward, the tools used here can often give misleading results. Namely, it has been shown that single variant eQTL lookups can lead to false positives, if GWAS loci harbour multiple independent eQTLs (Liu et al., 2019). While co-localisation analyses, such as SMR-HEIDI (Zhu et al., 2016) used within this work, have been specifically designed to use LD information to mitigate such issues, they can still fail in cases were independent eQTLs for separate genes are in complete LD (Zhu et al., 2016). Therefore, care needs to be taken when attempting causal gene assignment at a GWAS locus.

3.4.2 The association interval at the *PLA2G6* locus co-localises with regulators of 3 genes

Both SMR and TWMR highlighted *MAFF* and *PLA2G6* as the genes whose expression was most strongly affected by the top associated variants. Thus,

publicly available gene expression resources were used to identify tissues of high and low expression of these genes, and select them for the 5C experiment.

In humans, *PLA2G6* was most strongly expressed and MAFF was least strongly expressed in the testes (with RNA expression data from (She et al., 2009). Thus, testis cells were initially selected. The *Pla2g6* protein is primarily found in the seminiferous tubule cells in human histological sections (Uhlén et al., 2015) and given that no human cell lines of them exist, we opted to do the 5C on mouse-derived tissues. The second cell type, brain cortex cells, were selected as human mutations in *PLA2G6* have been linked to neurodegenerative and Parkinson-like diseases with iron accumulation in the brain (Morgan et al., 2006; Paisan-Ruiz et al., 2009). Given the extent of DNA and protein homology (as seen in Figure S4 and Table S3, accordingly), regulatory elements driving gene expression at this locus are likely to be shared between the two species.

For both of these tissues, there was observed enrichment for three sets of long range interactions. Namely, between the interval of association and the upstream regions of *Pick1*, *Dmc1* and *Sox10*. These interactions co-localised with areas that harboured many putative regulatory elements. Figure 3.3 highlights five types of regulatory signals within the assayed region. The first displayed track,

shows CTCF sites, which generally mark repression of gene transcription (Kim et al., 2015). All others are histone modifications that have been shown to be indicators of gene activation or repression. For the second track, H3K27me3 is able to recruit polycomb repressive complexes and thus silence gene expression (Ku et al., 2008). H3K4me1 marks poised chromatin regions that harbour enhancers, regardless of their activation state, while if coupled with H3K27me3 marks is more indicative of repression (Spicuglia and Vanhille, 2012, Zentner et al., 2011). Contrarily, H3K27a and H3K4me3 both mark active genomic elements; enhancers for the former and promoters for the latter (Creyghton et al., 2010, Heintzman et al., 2007). The ENCODE data for all of these marks were measured in 8-week old mice and should thus be highly comparable to the cells that were used for the 5C analysis.

In the context of this, the chromatin interaction with *Pick1* overlaps with very similar marks in both brain and testis cells. Namely, there were CTCF, H3K4m1, H3K27a and H3K4m3 peaks, indicating an active promoter and enhancer at this site. For the interaction upstream of *Dmc1*, there was a CTCF peak in the brain indicating repression of expression in this tissue, while H3K4m1 and H3K4m3 peaks in the testis cells indicate an active promoter and enhancer. Finally, the last

interaction with *Sox10*, contains CTCF repressive marks in the brain, but an H3K4me3 marked active promoter in the testis.

Evidence for the physical interaction of the association interval with these regions could mean that the genetic variation introduced by the top SNPs altered the binding between these regulatory elements. That would alter the expression patterns of the aforementioned genes. A change in binding could be further explored by doing a 3D fluorescence *in situ* hybridisation (3D-FISH) experiment in cells with targeted genetic alterations that recapitulate the SNP effects. Further work would be needed to understand how expressional changes in these genes could contribute to the GWAS phenotype. While this does not preclude short-range interactions, such as the ones towards *PLA2G6* or *MAFF*, it does provide the first evidence towards long-range regulation at this locus and demonstrates how complicated genome regulation can be.

3.5 Conclusion

In this chapter, we tried to identify the genes driving the Leg FMR phenotype at the two prioritised iDXA loci, nearby genes *ADAMTS14* and *PLA2G6*. We did so using both gene expression data and chromatin conformation studies. eQTL data

readily pinpointed *ADAMTS14* at one locus, while it highlighted many genes at the other locus. After conducting a 5C experiment we showed that even though the association interval at the latter locus was within the gene *PLA2G6*, which was a strong metabolic candidate (Section 2.4.2), the associated variants showed long-range interactions with three other genes, *Pick1*, *Dmc1* and *Sox10*. This, alongside the *FTO* example (Section 3.1.1), indicates why assigning causal genes at GWAS loci is important prior to any functional follow-up as the causal gene may not always be the closest one.

4

Adamts14, a novel candidate for body composition

4.1 Introduction

4.1.1 Metalloproteases in adipose tissue remodelling

As discussed in Section 1.4.2, nutrient excess causes AT to expand rapidly, in order to safely store excess energy as lipid in adipocytes, which is essential to maintain normal metabolic homeostasis. The ECM plays a vital role in accommodating the remodelling of AT in obesity. Firstly, the ECM helps maintain the structural integrity of the expanding cells and secondly, it is a source of mechanical and secreted signalling cues that aid growth and differentiation. Therefore, understanding the mechanisms behind ECM generation and degradation could provide important insights into AT remodelling.

The ECM of adipose is made up of several cellular components, mainly, collagens, laminins, fibronectin and proteoglycans. Collagens play a uniquely important role in AT remodelling, with deposition of collagen (types I–IV) being observed during adipose differentiation (Nakajima et al., 2002). There are several proteolytic enzymes that are able to dynamically modify all these components,

through ECM deposition and degradation. The two major protein families that are able to do so are the matrix metalloproteinases *(MMP)* and the a-disintegrin and metalloproteinase with or without thrombospondin motifs (*ADAMTS* and *ADAM*, respectively) (Nagase et al., 2006).

Such metalloproteinases can hydrolyse both matrix and non-matrix proteins and rely on the presence of zinc ions to do so. They exist in the

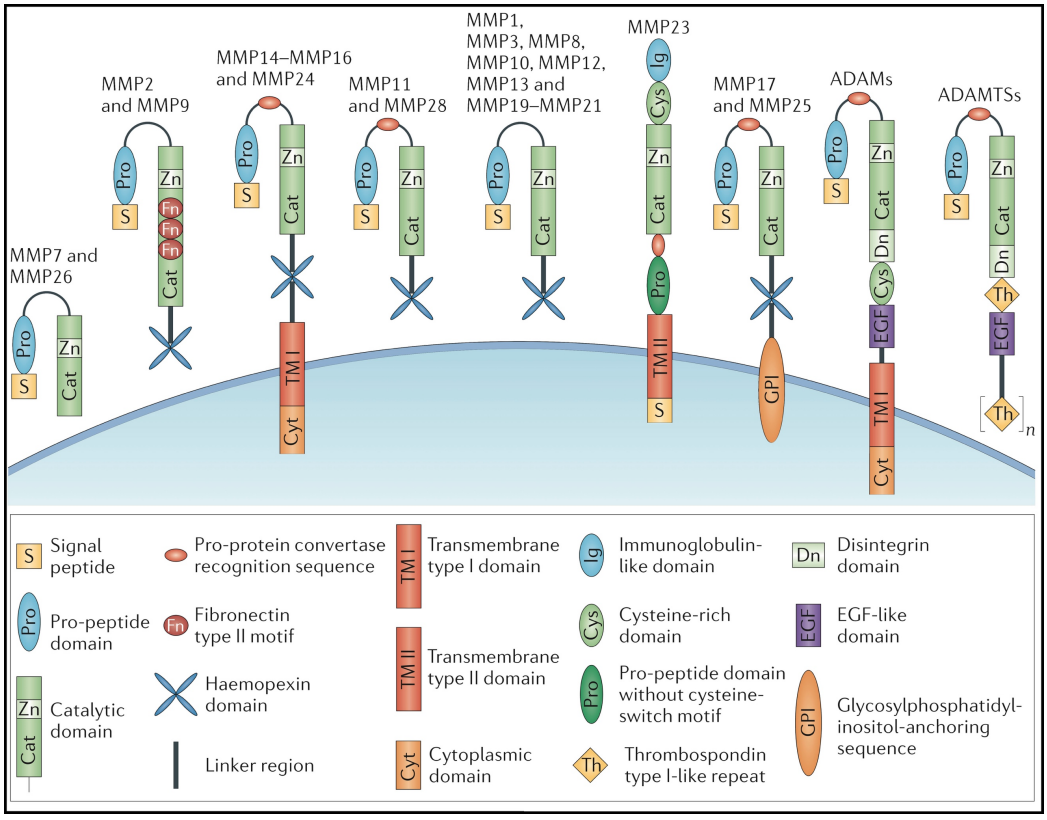


Figure 4.1 | The molecular structure of *MMP* and *ADAM(TS)* metalloproteinases.

MMPs are generally made up of four domains; the propeptide one, the catalytic one, the linker region and the haemopexin domain. The propeptide domain interacts with a zinc ion in the catalytic domain and ensures that the protein remains inactive until the propeptide region is removed. ADAMs proteins have a similar structure, but instead of the haemopexin, have additional domains that allow them to interact with the substrates from other cells. *Taken from (Khokha et al., 2013).*

extracellular environment of cells and have multiple domains with distinct functions (Figure 4.1). They can be soluble or anchored to the cell membrane and facilitate the delivery of cellular cues, through unique posttranslational modifications (Khokha et al., 2013).

In humans, there are 24 *MMP* and 22 *ADAM* genes and many studies, particularly those from mouse transgenic models, have shown the requirement for both protein families during AT expansion. For example, *Adam12*$^{-/-}$ animals show minor defects in BAT formation (Kurisaki et al., 2003), while muscle-specific over-expression of *Adam12* induces ectopic adipogenesis in their skeletal muscle (Kawaguchi et al., 2002). *Adamts1*$^{-/-}$ mice exhibit wide-spread developmental abnormalities and are leaner with less gonadal fat than control littermates (Shindo et al., 2000). Generalised pharmacological *MMP* inhibition suppresses adipogenesis in an *in vitro* and *in vivo* setting (Park and Yoon, 2012, Lijnen et al., 2002), suggesting the potential for therapeutic modulation of the ECM in obesity. Specific inhibition of *MMP2* can induce *in vitro* adipogenesis (Van Hul et al., 2013), while it has the opposite effect *in vivo* and *Mmp2*$^{-/-}$ mice show decreased weight gain and fat pad mass, with adipocyte hypotrophy, after 15 weeks of HFD (Van Hul and Lijnen, 2008). *Mmp19*$^{-/-}$ mice show increased weight gain and overt adipocyte

hypertrophy, after a 30-week HFD administration (Pendás et al., 2004). Finally, *Mmp14*$^{-/-}$ mice exhibit a general decrease in AT mass, notably preadipocytes from these mice differentiate normally under 2D *in vitro* culture conditions, while if cultured in a 3D collagen matrix, the *Mmp14*$^{-/-}$ preadipocytes fail to differentiate into mature adipocytes (Chun et al., 2006, Chun et al., 2010).

4.1.2 Limited evidence for the function of *Adamts14*

Adamts14 belongs to the *ADAMTS* family of metalloproteinases, was first discovered in 2002 and is highly homologous to *Adamts2* and *3* (Colige et al., 2002). All three have the ability to cleave collagens at the aminoterminal propeptides and thus facilitate the maturation of fibrillar collagen (Bekhouche et al., 2016). *Adamts2*$^{-/-}$ mice exhibit skin fragility due to accumulation of type I collagen propeptides, while *Adamts3*$^{-/-}$ mutations are lethal at E15 due to the inability to cleave (non-collagenous) growth factors that are needed for correct lymphangiogenesis (Janssen et al., 2016, LI et al., 2001). This indicates that the functionality of these proteins goes beyond collagen processing.

Adamts14 has not been studied at an equivalent depth. Recently, *Adamts14*$^{-/-}$ mice were created and reportedly showed no observable phenotype (Dupont et al., 2018). Dupont and colleagues showed that maturation of type I collagen was

unaffected in the skin, tendon and bone of *Adamts14*$^{-/-}$ animals. *Adamts2*$^{-/-}$ *Adamts14*$^{-/-}$ double-null animals showed comparable phenotypes to the single mutants, suggesting *Adamts14* was dispensable during *in vivo* procollagen processing, due to functional redundancy between the two homologues. However, these observations were restricted to the few tested tissues and thus do not preclude a more tissue-specific role for *Adamts14* in collagen processing.

Of note, Dupont et al. unexpectedly observed dermatitis-like skin lesions form on the skin of the double mutant mice, that were unrelated to the aforementioned skin fragility phenotype. They concluded that *Adamts2* and *14* act as regulators of immune activation and that their simultaneous deficiency triggered an intrinsic skin immune response. This further corroborates the functional versatility of these proteins.

4.1.3 Hypothesis & Aims

After providing eQTL evidence that *ADAMTS14* was the likely causal gene at this association interval, we hypothesise that *Adamts14* was a novel modulator of adipogenesis. To test this, mice lacking the *Adamts14* gene were obtained and subjected to metabolic assessment, prior to and after HFD exposure, with the following aims:

I. To assess whether the reduced adiposity phenotype seen in humans carrying the *ADAMTS14* variants was also evident in mice lacking one or both alleles of *Adamts14.*

II. To investigate whether loss of *Adamts14* causes an adiposity and/or metabolic phenotype.

III. To start investigating the cellular mechanism underlying any effect of *Adamts14* deficiency.

4.2 Experimental design & procedure

4.2.1 Mouse procurement & husbandry

The *Adamts14*$^{-/-}$ mice were created as described in (Dupont et al., 2018) and shown in Figure S7 and obtained in collaboration with the Colige group at the University of Liege. Sperm from heterozygous C57BL/6 animals was shipped to Edinburgh frozen and used to perform *in vitro* fertilisation (IVF) to C57BL/6 females, at the BVS Transgenic Core facility. Resulting heterozygous animals were initially metabolically characterised and also bred to create colonies of *Adamts14*$^{-/-}$ and *Adamts14*$^{+/+}$ animals.

All experiments were performed under PPL 60/8117, appropriate PILs granted under the Home Office Scientific Procedures (Animals) Act 1983 and after full ethical review by the University of Edinburgh Biological Sciences Services. Male mice were used for all experiments and were maintained single-housed in either standard or individually ventilated cages with *ad libitum* access to food and water at the Little France BRR facility. They were kept at 19–22°C and maintained with a 12-hour light/dark cycle with lights on at 7am.

4.2.2 Genotyping

Genomic DNA from all live-born mice was extracted from ear tissues and used to genotype by PCR for the targeted alterations. DNA extraction was performed via the digestion of the ear tissues in 20 µl DNAreleasy (1:3 dilution in water, Anachem, UK) at 75°C for 5 minutes, followed by 95°C for 2 minutes. 1–2 µl of DNA were used as a template per PCR reaction. PCR reactions that specifically amplify the targeted alteration were designed. For each mouse, two reactions were multiplexed, one that amplifies a fragment within the inserted cassette that induces the KO and one that amplifies the region that would be disrupted by the insertion. This allows for the differentiation between all three possible genotypes;

WT, heterozygous and homozygous KO. Cycling protocols and primer combinations can be found in Table S6 in the Appendix.

4.2.3 *In vivo* metabolic phenotyping

4.2.3.1 Diet & weight monitoring

At the age of approximately two months, the mice were subject to a 58% high fat diet (Research Diets Inc, 12331) for up to 15 weeks by allowing the animals ad libitum access to the diet. Animals were weighed weekly and had their lean and fat mass determined by time domain–nuclear magnetic resonance (TD-NMR, Bruker) at multiple time points throughout each experiment (shown in Figure S8).

4.2.3.2 Indirect calorimetry

PhenoMaster cages (TSE systems, Germany) were used to assess energy expenditure, locomotor activity and food and drink intake of the mice at the beginning and end of the experimental period. To do this, mice were placed in individual monitored cages with *ad libitum* access to food and water. The animals were allowed to acclimatise to their environment for 24 hours before collection of experimental data. Measurements for each parameter were taken continuously for the period of at least 24 hours and were recorded every 15 minutes.

4.2.3.3 Oral glucose tolerance tests (oGTTs)

oGTTs were carried out at least at the beginning and end of the HFD treatments. On the morning of the oGTT, the animals were fasted for 5 hours by removing their food and transferring them to fresh cages, with *ad libitum* access to water. A 20% glucose solution in water was ingested by oral gavage at a concentration of 2 mg glucose per gram of body weight. Blood glucose levels were measured using a glucometer (OneTouch Ultra, LifeScan), prior to glucose administration and after at 15, 30, 60 and 120 minute intervals from a small drop of blood from a tail-nick.

4.2.4 Tissue collection & analysis

At the end of experiments, animals were sacrificed humanely; tissues were harvested and immediately frozen under dry ice and stored at -80°C or collected in 4% paraformaldehyde (PFA) and then dehydrated for 24 hours in 75% ethanol.

Fixed and dehydrated tissues were embedded in paraffin and sectioned. For collagen quantification, 5 µm sections were stained with Picro Sirius Red (PSR) (Abcam, ab150681). Briefly, slides were baked at 55°C for an hour, dewaxed in xylene and rehydrated across an ethanol gradient. Rehydrated sections were then re-fixed in PFA for an hour and then stained with PSR for another hour. Stained

sections were rinsed quickly in two changes of 0.5% acetic acid solution and then dehydrated across an ethanol gradient and cleaned with xylene. Mounted sections were viewed with brightfield microscopy. Histomorphometry of adipocytes from the gonadal and gluteofemoral depots was assessed using the Adiposoft plugin in ImageJ.

4.2.5 Statistical analysis

All statistical analyses were carried out in R, statistical programming language, by Two-Sample t-test or a Two-way ANOVA (and using body weight as a covariate in the case of EE in Figure 4.4), as indicated in the relevant figure legends. Where indicated, post-hoc analyses were conducted using a Tukey test.

4.3 Results

4.3.1 Mice heterozygous for the deletion of *Adamts14* show reduced weight gain and adiposity and improved glucose tolerance upon HFD exposure

Initially, due to a breeding delay, a cohort of 3-month-old heterozygous (*Adamts14*$^{+/-}$) and WT (*Adamts14*$^{+/+}$) male mice were exposed to a HFD for 13 weeks (n=5 per group), ostensibly to assess potential dosage effects, once *Adamts14*$^{-/-}$ mice were obtained. The *Adamts14*$^{+/-}$ mice were more resistant to weight gain than their WT littermates over the 13-week period ($p= 8.3x10^{-15}$, Figure 4.2A). This was

also reflected in how they gained less fat (p=0.01) and lean mass (p=0.04) over this time (Figure 4.2B), especially at the 6-week timepoint. However, this difference was attenuated with the longer exposure to the diet. Indeed, *post mortem* dissection of the adipose depots after the 13-week period showed no genotype difference for fat mass (Figure 4.2C). Notably, there was also no significant difference in any of the body size indices tested (Figure 4.2B-inset).

Consistent with their relative resistance to fat mass gain *Adamts14*$^{+/-}$ mice exhibited comparable baseline glucose tolerance, but significantly better glucose tolerance at 6 weeks post-HFD (p=0.01), an effect that was partially maintained but lessened after 13 weeks (Figure 4.2D). This was supported by post-hoc analyses, which showed a marked difference in glucose tolerance of both the heterozygous and WT cohorts between oGTT1 and 2 ($p=3.4x10^{-3}$ and $p<2.2x10^{-16}$, accordingly for the two genotypes), but not 2 and 3 (p=0.8 and p=0.5).

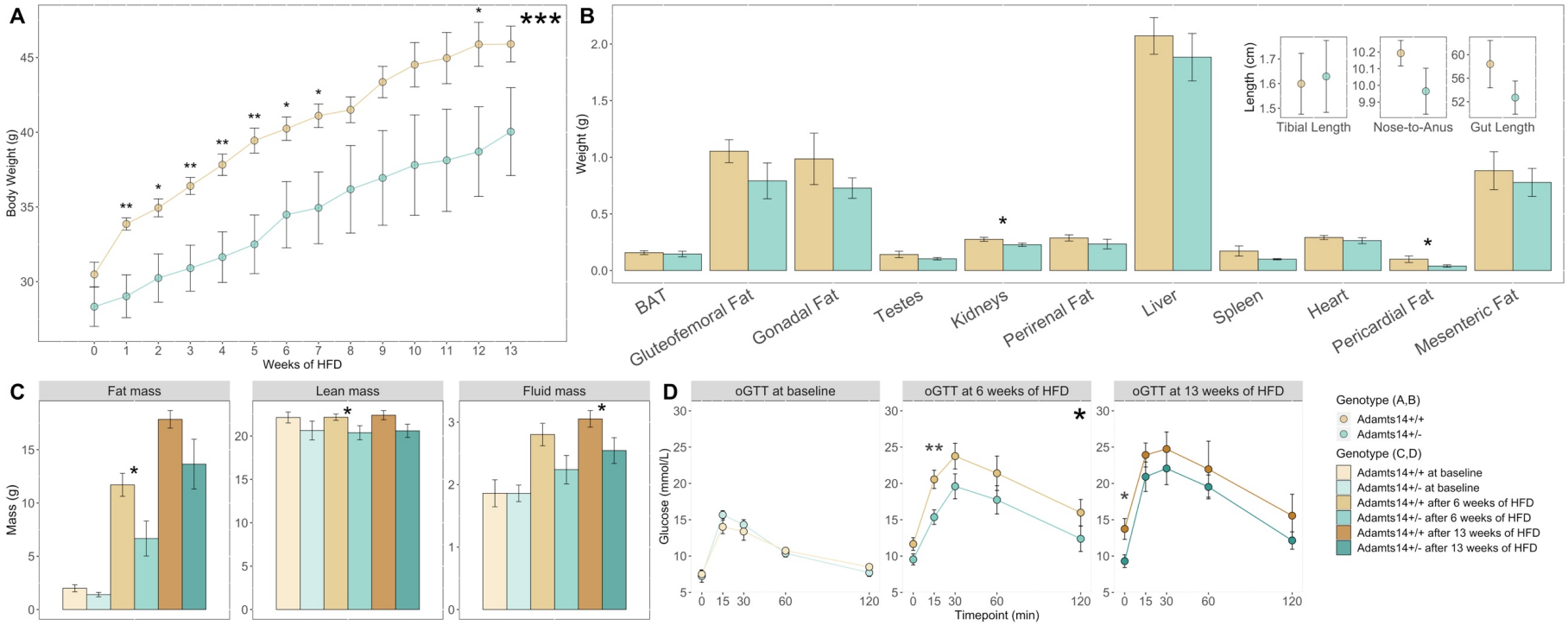


Figure 4.2 | The *Adamts14*$^{+/-}$ mouse was resistant to weight and fat gain and had improved glucose homeostasis under HFD conditions.

(A) Body weight of WT and heterozygous animals through 13 weeks of HFD, starting when the animals were 3 months old. (B) Post mortem organ and tissue weights of the WT and *Adamts14*$^{+/-}$ mice. (B-inset) Comparison of length of the left tibia, the mouse from nose to anus and the gut. (C) Fat, lean and fluid mass changes before HFD and after 6 and 13 weeks. (D) Oral glucose tolerance tests, at 0, 6 and 13 weeks of HFD. N=5 per genotype. Data expressed as mean ± se. T-tests done for direct comparisons in A, B, C and D and a 2-way ANOVA done over the whole experimental period of A and D. Significance denoted as * for p<0.05, ** for p<0.005 and *** for p<0.0005.

4.3.2 *Adamts14*$^{-/-}$ mice show reduced weight gain under HFD and have increased energy expenditure, without overt changes in overall adiposity

After establishing a homozygous cohort, we sought to metabolically characterise them at the same level as the heterozygous cohort. *Adamts14*$^{-/-}$ animals were viable, bred normally and did not exhibit embryonic lethality. This cohort consisted 8 males per genotype (*Adamts14*$^{+/+}$ and *Adamts14*$^{-/-}$), aged around 2 months. Null mice also showed reduced weight gain with exposure to HFD ($p=9x10^{-8}$, Figure 4.3A). Glucose tolerance via oGTT and body composition by TD-NMR were assessed at the beginning and end of the HFD administration (0 and 13 weeks) and no significant differences were found (Figure 4.3A&C). No overt differences were shown *post mortem* for organ weights (Figure 4.3B), however the *Adamts14*$^{-/-}$ animals had significantly smaller tibial and gut lengths, indicating a potentially smaller body size (Figure 4.3B-inset).

Mice were also assessed in terms of their energy expenditure (EE), activity and food intake before and after the administration of the HFD (Figure 4.4 and Figure 4.5). The Adamts14-/- animals consumed more food compared to their *Adamts14*$^{+/+}$ littermates, both on normal diet and HFD at the end of the experiment (Figure 4.4A&E, Figure 4.5A). They also exhibited higher EE at both of these

timepoints (Figure 4.4D&H). The physical activity of the *Adamts14*$^{-/-}$ animals was higher initially, but was the same as the *Adamts14*$^{+/+}$ mice after the 13 weeks of HFD administration (Figure 4.4B&F). The opposite was seen for their respiratory exchange rate (RER), which is an index of carbohydrate versus lipid fuel utilisation. Namely, both groups exhibited comparable levels at the beginning of the experiment, but the *Adamts14*$^{-/-}$ mice had significantly higher RER at the end of the experiment (Figure 4.4C&G).

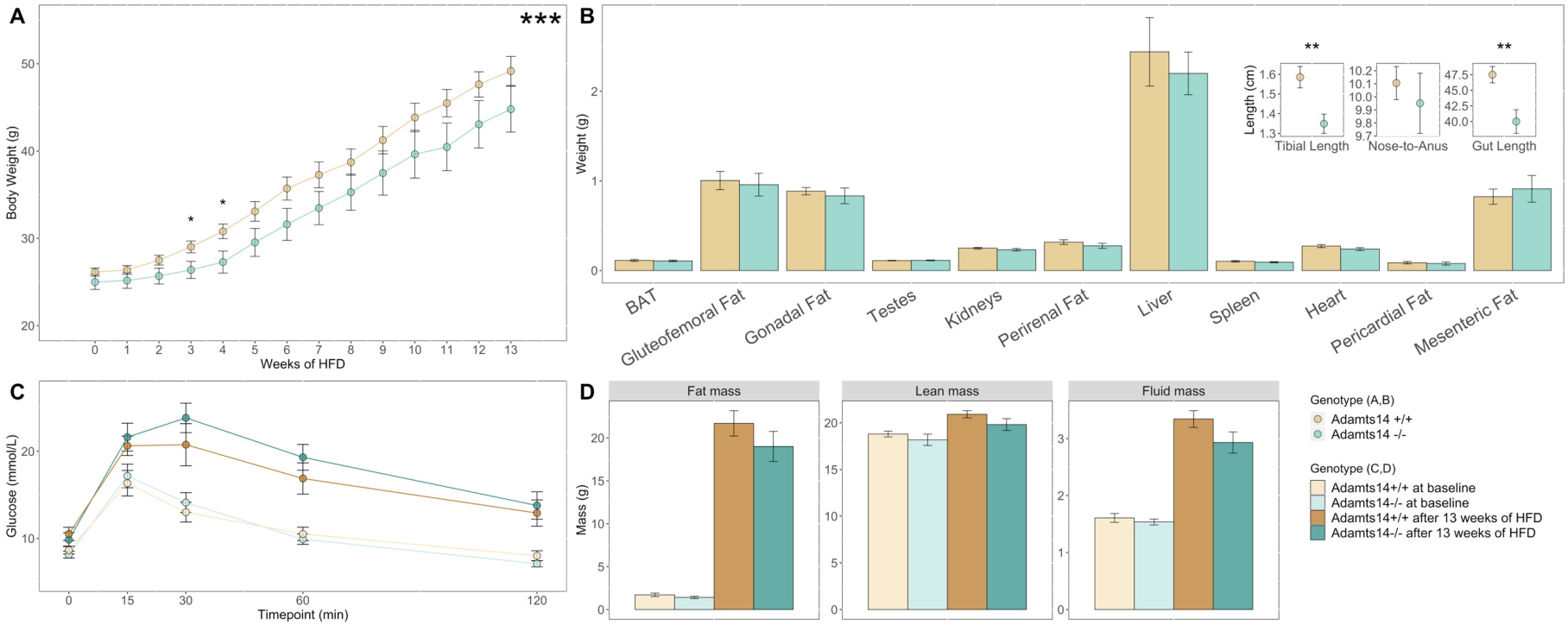


Figure 4.3 | The *Adamts14*$^{-/-}$ mouse was also more resistant to weight gain under HFD conditions, but shows no differences in adiposity.

(A) Body weight of WT and KO animals through 13 weeks of HFD. (B) Post mortem organ and tissue weights of the WT and *Adamts14*$^{-/-}$ mice. (B-inset) Comparison of length of the left tibia, the mouse from to anus and the gut. (C) Fat, lean and fluid mass changes before and after 13 weeks of HFD. (D) Oral glucose tolerance tests, at before and after the HFD exposure. N=8 per genotype initially, down to 6 WT and 7 KO by 13 weeks. Data expressed as mean ± se. T-tests done for direct comparisons in A, B, C and D and a 2-way ANOVA done over the whole experimental period of A and C. Significance denoted as * for p<0.05, ** for p<0.005 and *** for p<0.0005.

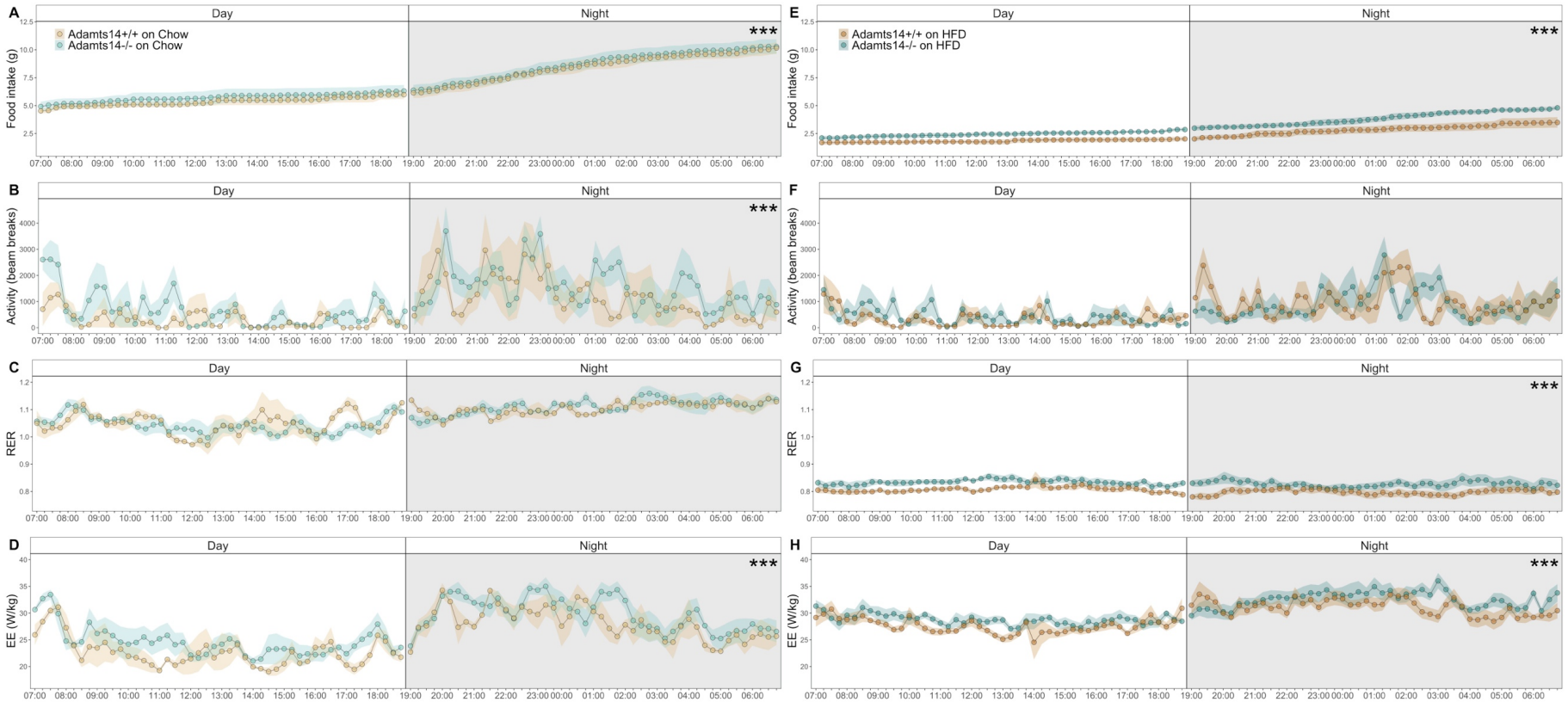


Figure 4.4 | The *Adamts14*$^{-/-}$ exhibits altered energy homeostasis and eating behaviour under both normal Chow diet and HFD conditions.

(A & E) Food intake (B & F) Activity (C & G) RER (D & H) EE all at baseline and after 13 weeks of HFD. For the Chow analyses, N was 5 KO and 3 WT (A–D) and 7 KO and 6 WT for the HFD analyses (E–H). Data expressed as mean ± se and was analysed by a 2-way ANOVA. Significance denoted as * for $p<0.05$, ** for $p<0.005$ and *** for $p<0.0005$.

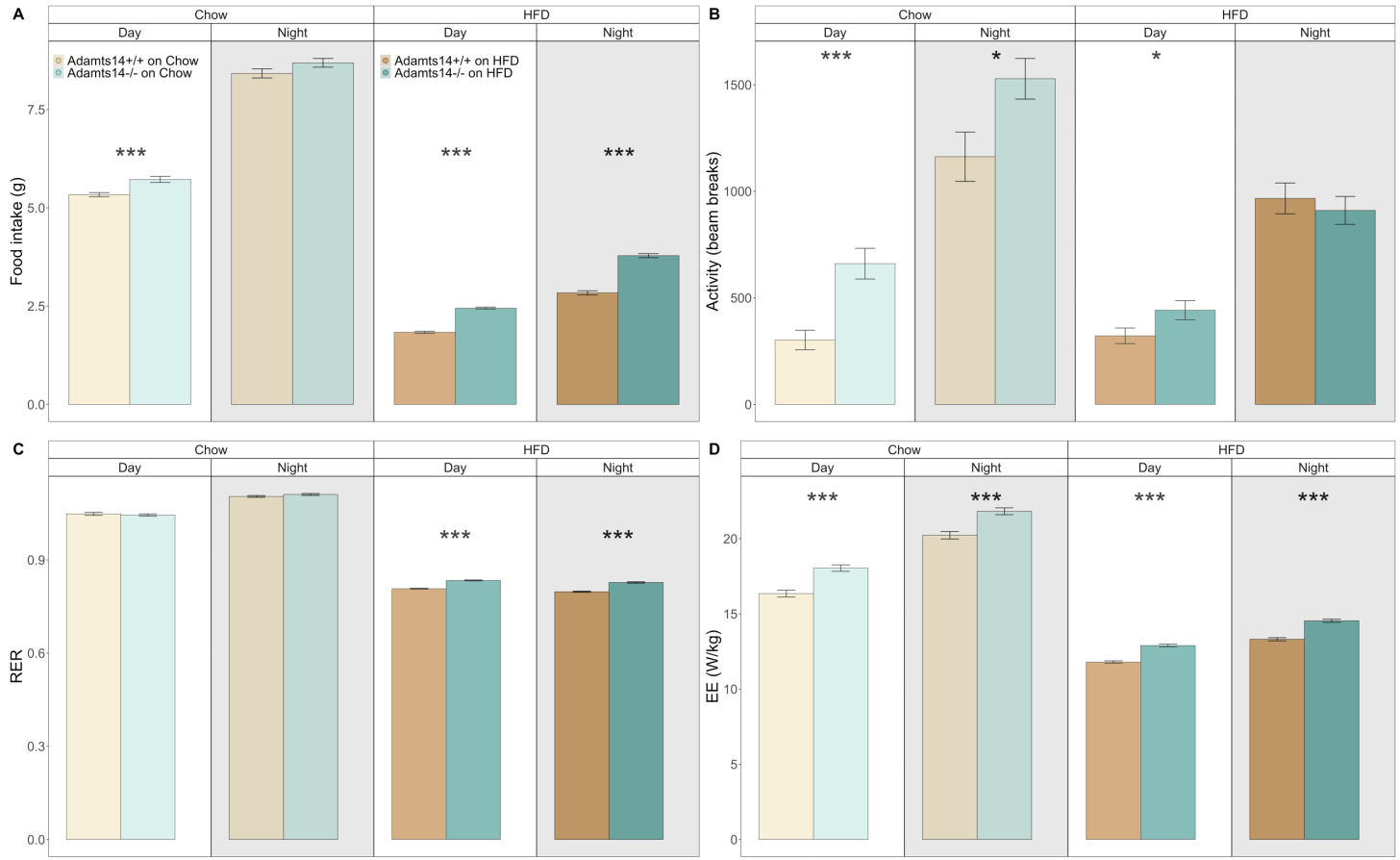


Figure 4.5 | The Adamts14-/- mouse homeostatic phenotype over light and dark phases.

(A) Food intake (B) Activity (C) RER and (D) EE corrected for body weight, from Figure 4.4, averaged for the 12-hour light and dark phases of the queried time period. For the Chow analyses, N was 5 KO and 3 WT and 7 KO and 6 WT for the HFD analyses. Data is presented as mean ± se and was analysed by a two-sided T-test. Significance denoted as * for p<0.05, ** for p<0.005 and *** for p<0.0005.

4.3.3 *Adamts14*$^{+/-}$ mice show no changes in adipocyte morphology

Post mortem tissues were collected (Figure 4.2B) and further analysed. Gross morphology of a selection of them revealed no differences between the *Adamts14*$^{+/-}$ and *Adamts14*$^{+/+}$ animals (Figure 4.6A–E). Quantification of adipocyte cell area and number in the gluteofemoral and gonadal depots, showed that heterozygous and WT animals had comparably sized cells in both depots (Figure

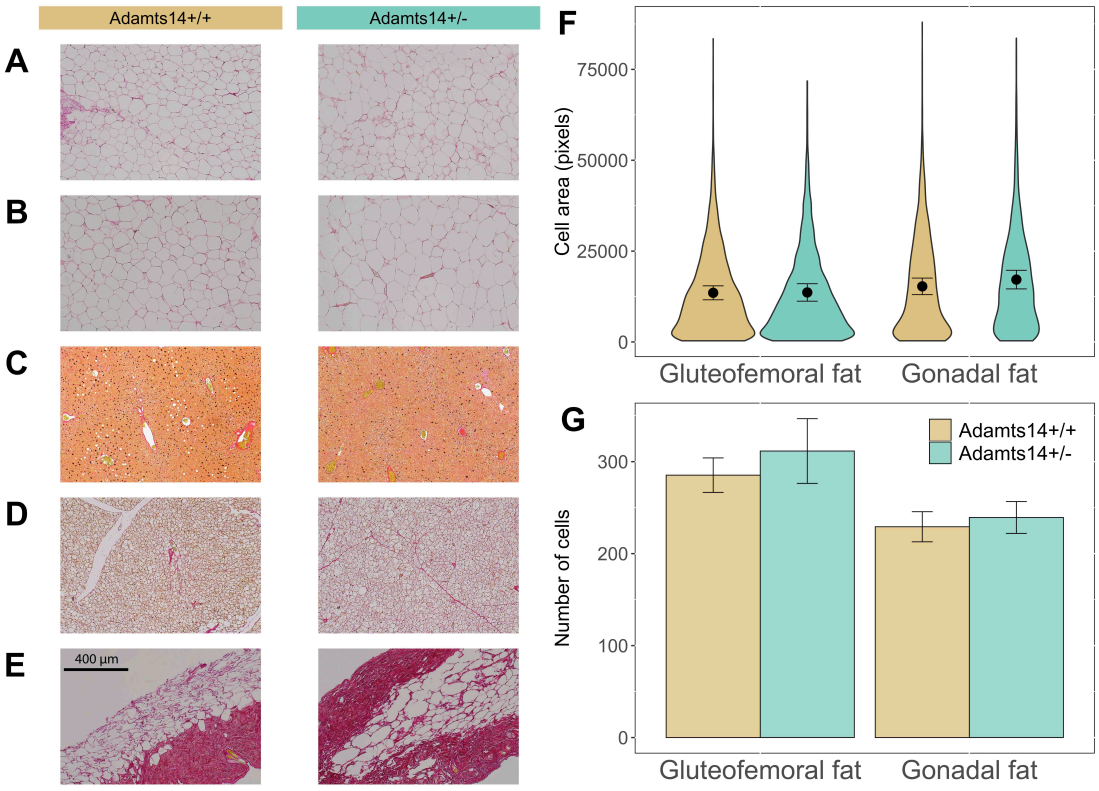


Figure 4.6 | Histomorphological analyses of tissues from *Adamts14*$^{+/-}$ and *Adamts14*$^{+/+}$ mice show no overt differences.

(A–E) Sections of gluteofemoral fat, gonadal fat, liver, BAT, and back skin, accordingly, stained with PSR. Scale bar given in E. (F) Distributions of adipocyte cell size in the gluteofemoral and gonadal depots, for heterozygous and WT animals. (G) Overall number of cells found per depot and genotype. N=3 animals per genotype and 5 independent images quantified per depot per animal. Data expressed as mean ± se and analysed by a two-sided t-test.

4.6F) and also similar numbers of cells (Figure 4.6G). This analysis would have been repeated for the homozygous animals, but had to be postponed due to Covid-19 work restrictions.

4.4 Discussion

In chapter 3, variants upstream of the gene *ADAMTS14* were associated with an increase of Leg FMR. The associated SNPs were eQTLs for tissue mRNA levels of *ADAMTS14* and indicated that decreased expression of human *ADAMTS14* caused a decrease in leg fat mass. In this chapter, we modelled the effects of the *Adamts14* gene deletion and haploinsufficiency on body composition and metabolic homeostasis, using the *Adamts14*$^{-/-}$ mouse.

4.4.1 The *Adamts14*$^{-/-}$ mouse phenotype is concordant with the iDXA GWAS

Adamts14 was identified as a putative driver of human body composition and it was therefore followed up in a murine model with a targeted deletion, that rendered *Adamts14* functionally null (Dupont et al., 2018). We observed that both heterozygous (*Adamts14*$^{+/-}$) and homozygous (*Adamts14*$^{-/-}$) animals for this mutation bred and developed normally and investigated the effect of the

mutations on body composition and energy homeostasis under HFD conditions, as outlined in Figure S8.

In terms of their body composition, both *Adamts14*$^{+/-}$ and *Adamts14*$^{-/-}$ animals were more resistant to weight gain throughout the HFD exposure, when compared to the WT mice. There were no significant differences between genotypes in total fat and lean mass at 0 and 13 weeks, however both were significantly lower in the *Adamts14*$^{+/-}$ animals at the 6-week timepoint. The same was seen for circulating blood glucose levels during the oGTTs and no difference was seen in *post mortem* fat pads weights in either cohort. This indicates that any protective effect of the genotype was masked by prolonged exposure to the diet. Exposing a cohort of *Adamts14*$^{-/-}$ mice to a shorter HFD treatment could produce concordant results and delineate any potential gene dosage effects.

Interestingly, the *post mortem* showed that the *Adamts14*$^{-/-}$ animals were smaller than the *Adamts14*$^{+/+}$ controls, at least in terms of their tibial and gut lengths, but not when measuring their nasoanal length. As the latter is the most commonly used measure of murine body size (Dickinson et al., 2016) this might be an indication of a more specialised phenotype, i.e. a bone or cartilage phenotype. This is intriguing given the incidental finding that some of the *Adamts14*$^{-/-}$ male

animals exhibit a short airorhynchic snout and reduced stature with variable penetrance (Figure S9) and also the recently reported bone density phenotype in a different strain of *Adamts14*$^{-/-}$ mice (Table S3). Alternatively, gut length is a common indicator of nutrient absorption and can fluctuate under different dietary qualities and availabilities (expanded in Section 5.2.2 and reviewed in (Hill Jr et al., 1968, Mortensen et al., 2010, Li et al., 2017)). No gross body size difference was found in the *Adamts14*$^{+/-}$ animals.

4.4.2 The effects of *Adamts14*$^{-/-}$ on energy homeostasis

At the time of the second indirect calorimetry experiment, the Adamts14$^{-/-}$ mice weighed, on average, 4.5 grams less than their WT littermates and within the queried 24 hours consumed, on average, 1.3 grams more food (Figure 4.3, Figure 4.4, Figure 4.5). They also expended more energy when compared WTs, both before and after the HFD administration. The higher EE could initially be explained by the higher physical activity seen prior to the HFD. However, physical activity was the same for the two genotypes after the 13 weeks of HFD, while EE remained relatively elevated in the KOs. However, whilst both genotypes displayed a flattening of RER, indicative of loss of metabolic flexibility and increased lipid metabolism after the HFD exposure, RER after the 13 weeks remained higher in

the *Adamts14*$^{-/-}$ animals. This could be merely a consequence of their hyperphagia, i.e. they consumed and oxidised proportionally more carbohydrates or, alternatively, due to maintenance of an intrinsically greater reliance on carbohydrate oxidation.

Taken together the data indicates an altered energy homeostasis in the *Adamts14*$^{-/-}$ mice and more specifically and increase in metabolic rate. This could be explained by an AT phenotype, such as impaired lipid storage or a BAT phenotype, such as increased thermogenesis. When looking at the histology of the Adamts14$^{+/-}$ mice, no phenotype was apparent in their BAT and their adipocytes had comparable morphology (Figure 4.6). The same analysis coupled with further metabolic studies of BAT or WAT oxidation in the *Adamts14*$^{-/-}$ mice are needed to determine the source of their altered energy homeostasis (see Section 5.2.2).

4.5 Conclusion

To summarise, we modelled the deletion of a GWAS candidate gene, *ADAMTS14*, in mice, in order to support its role as a novel driver of body composition. In the human GWAS data, variants that associated with decreased expression of *ADAMTS14* caused a decrease in leg fat mass. We showed that the

in vivo data corroborates these findings and mice with mutations in the *Adamts14* gene showed reduced weight gain and adiposity under a short course of HFD. *Adamts14*$^{-/-}$ mice also exhibit hyperphagia and increased energy expenditure, which indicates an increase in their metabolic rate. More work is needed to determine the molecular basis of this phenotype.

5

Discussion

5.1 Summary of key findings

In this thesis, I firstly developed a method to simply and efficiently predict DXA phenotypes in UKB. To create a prediction model, I used anthropometric and BIA proxies of body composition in the 5,000 participants that also had DXA measured (UKB_{DXA}) and used this model to predict them in the rest of the cohort (UKB_{iDXA}). The iDXA phenotypes provided with accurate predictions of the DXA ones (Pearson's R=0.81, overall). They also correlated well with obesity disease sequelae; iDXA phenotypes predictive of android patterns of AT distribution showed positive correlations with CVD and T2D, while those indicative of gynoid patterns did not.

I then performed GWAS on the iDXA phenotypes in the UKB_{iDXA} cohort. Exclusion of related and non-white-British individuals, brought the UKB_{iDXA} cohort size to approximately 390,000 participants. After correcting for multiple testing, I mapped the GWS signals across the iDXA traits to over one thousand

independent loci. The majority of these were known obesity associations and after filtering for them, there were 242 novel adiposity associations.

I sought to replicate these loci in an independent MA cohort of the equivalent DXA phenotypes. The collective size of this cohort was 18,000 participants and, thus, only SNPs with sufficiently large effect sizes to be discoverable in this much smaller cohort were sent for replication. Out of the 27 submitted SNPs, 6 replicated at an FDR of 10%, while 19 were exerted directionally consistent effects between the DXA and iDXA. These included SNPs in or around genes *PLA2G6, CPS1, ACVR2B, FBXO36/TRIP12, ADAMTS14* and *DLG4/ACADVL*.

I prioritised the *PLA2G6* and *ADAMTS14* loci for experimental follow-up for two quite different reasons. The *PLA2G6* locus had been previously nominally associated with body fat percentage (Lu et al., 2016) and also blood lipid levels (Willer et al., 2013), but was a gene-rich region and no attempts had been made to discern a causal gene at the locus or to understand how it affected body composition. In contrast, almost nothing was known about the *ADAMTS14* locus, but the region was less gene dense and other members of the *MMP* and *ADAM(TS)* families of genes play well-established roles in development and body composition. Thus, work on both of these loci could provide novel insights into the

genetics of human adiposity. For *PLA2G6*, further dissection of the locus might help pinpoint a causal gene behind a robust association to various facets of metabolism. For *ADAMTS14*, it could provide with the opportunity to identify and validate a novel gene that affects body composition.

Using *in silico* tools for causal gene assignment, suggested that SNPs within the interval of association at the *PLA2G6* locus causally affected the expression of many genes in many tissues. Thus, we used 5C to look at chromatin dynamics and interactions between the top associated SNPs and nearby genomic elements. In doing so, we observed long-range interactions between the associated interval and the upstream regions of genes *Pick1*, *Dmc1* and *Sox10*, which coincided with regulators of these genes. These interactions could be further corroborated in a 3D-FISH setting in future work.

In contrast to the *PLA2G6* locus, *ADAMTS14* was strongly supported as the causal gene driving the association signal. Thus, we modelled the effects of null mutations in a murine system. The mouse model phenocopied the human GWAS data; the SNPs associated with a decrease in gene expression and also fat mass, while the *Adamts14*$^{-/-}$ and *Adamts14*$^{+/-}$ mice weighed less after exposure to a HFD. Moreover, mice heterozygous for the *Adamts14* null mutation exhibited lower fat

mass and improved glucose homeostasis after an initial 6 weeks of HFD. This protective effect became less pronounced with chronic exposure (13 weeks) to the HFD. Indirect calorimetry on the homozygous animals showed that they were hyperphagic and had a higher energy expenditure and RER, both initially on control diet and after the administration of the HFD, indicating an altered energy homeostasis. Histology on tissues from the heterozygous animals did not show any obvious adipogenic defects, as evidenced by adipocyte cell size or number, that might help explain the phenotype in the homozygous animals. Clearly, more work is needed to understand the ways by which *Adamts14* exerts its effects. However, this work clearly implicates it as a novel driver of body composition and metabolism.

5.2 Limitations & future work

The main aim of this work was to identify novel gene drivers for human adiposity phenotypes and to discover their function. Firstly, by imputing body composition phenotypes using proxies and further by integrating human and mouse biology. As a direct consequence of this methodology, this study has several limitations.

5.2.1 The iDXA GWAS

Whole body imaging is the only non-invasive way to accurately quantify human adiposity. In the context of GWAS, the largest GWAMA of BMI includes 35 times as many participants as the largest DXA one (Yengo et al., 2018, Neville et al., 2019). To try to ameliorate this situation, I predicted whole body imaging data in UKB and did so with 80% overall accuracy. This resulted in an effective iDXA sample size of 312,000 (i.e. 80% of the total iDXA GWAS cohort size), which is more than an order of magnitude greater than the sample size available for the direct DXA measures alone (UKB_{DXA} n=4,400). This led to an increase in discovery power to identify novel genes that associate with the DXA adiposity phenotypes.

The anthropometric and BIA components I used in the imputation models were predictive over the DXA phenotypes, as both infer body composition. However, it remains unclear what proportion of the GWAS loci were DXA signals or signals for the BIA and anthropometric proxies used in the prediction models. As mentioned in Section 2.4.1, the only way to directly test this would be to do so in an equally sized and powered DXA cohort, which is impossible for the time being.

After the iDXA GWAS work had taken place, Johansson and colleagues conducted GWAS on the same cohort, focusing on BIA-based segmental adiposity (Rask-Andersen et al., 2019). In doing so they identified 29 novel associations, many of which were highlighted in my own work and included many metalloproteases, *ADAMTS14* inclusive. Arguably, since both studies were conducted in the same cohort and phenotypes and we observed a near 10-fold increase in the number of novel associations, this could be an indication that the iDXA phenotypes captured closer representations of body composition than the BIA proxies alone could.

Shortly thereafter, Johansson and colleagues published again, this time demonstrating that DXA-measured VAT can be predicted using anthropometric and BIA phenotypes in UKB (Karlsson et al., 2019). Their predicted VAT GWAS yielded 102 novel loci and was further used to demonstrate causal relationships between visceral adiposity and diseases, such as CVD, T2D and hyperlipidaemia. This had never been shown to such a scale, due to the unavailability of VAT-measures in large cohorts. They also showed a striking sex difference for risk of developing T2D, with females carrying a 3-fold increased risk (odds ratio (OR) of 7.3 in women and 2.5 in men). The same trend was shown in a study looking at the

causal relationship between BMI and T2D (Censin et al., 2019), however the reported ORs for this study (3.8 in women and 2.8 in men) were half of the ones reported by Karlsson et al.. These two studies are not directly comparable as the MR instruments in Censin et al. were constructed with data from the GIANT-UKB MA, but the comparison gives an indication that predicted VAT mass serves a far better measure of central adiposity than BMI, at least in a disease causality context. This corroborates the argument that the DXA phenotype imputation used within this thesis has the potential of delivering novel adiposity genetics insights, outwith the genetic drivers of anthropometric and BIA phenotypes.

On the matter of sexual dimorphism in body composition, the above studies showed that increased adiposity causes a higher risk of downstream disease in women than in men (Censin et al., 2019, Karlsson et al., 2019). In terms of obesity GWAS, associated variants often exhibited larger effects in women-only analyses (Pulit et al., 2017, Justice et al., 2019, Rask-Andersen et al., 2019), while some had no effects at all on male body composition (Fox et al., 2012, Heid et al., 2010). This all indicates that in the context of adiposity and its genetic drivers, a lot of the biological mechanisms appear to be sex-specific. Indeed, human AT distribution patterns are highly sex-dimorphic; women carry more AT overall and accumulate

fat on their lower halves, while men carry it on their upper body (Pasco et al., 2014).

Thus, a limitation of this work is that it was conducted in a manner that is semi-agnostic to sex. While sexual dimorphism was taken into account when predicting the iDXA phenotypes, as all models were fitted separately for the two sexes, the iDXA GWAS did not account for sex differences, as the ones seen above. In fact, the way they were conducted was ideally suited to SNPs that exerted equally sized effects on the body composition of men and women. Indeed, for the case of the top SNP at the *ADAMTS14* locus, rs12359330, the effect was near equal for men and women (β=0.02±0.003). Thus, it is possible that the iDXA GWAS could have masked loci with SNPs that exert highly sex-specific effects on adiposity.

5.2.2 Energy homeostasis & the *Adamts14*$^{-/-}$ mouse

Not having to account for sex differences made experimental follow-up simpler. When conducting the *in vivo* study, we only conducted experiments on male mice. While there was no genetic reason to expect the female mouse study results will be different, it is acknowledged that female cohorts will be important

to complete an understanding of the role of *Adamts14* in adiposity and metabolic disease, when the opportunity presents.

More work is needed in order to explain the increased EE phenotype we observed in the *Adamts14*$^{-/-}$ mice. WAT and BAT histology on the *Adamts14*$^{-/-}$ *post mortem* tissues, might show structural differences compared to the WT tissues and begin to offer an explanation for the biological basis underlying their divergent homeostasis. Other *MMP/ADAM(TS)* gene family members have also been implicated in WAT and/or BAT development, as discussed in Section 4.1.1. For *Adam12*, 30% of *Adam12*$^{-/-}$ mice exhibit impaired BAT formation and die prior to weaning (Kurisaki et al., 2003), while an overexpressing mouse line shows ectopic intramuscular adipogenesis (Kawaguchi et al., 2002). This indicates that *Adam12* acts as a regulator of both myogenesis and adipogenesis (Kurisaki et al., 2003). Equivalently for the *Adamts14*$^{-/-}$ mice, increased thermogenic activity in the BAT and impaired adipogenic WAT capacity would explain the observed increase in EE and reduced weight gain.

Impaired adipogenic potential was also seen in *Mmp14*$^{-/-}$ mouse-derived preadipocytes (Chun et al., 2006) and also in mice that underwent pharmacological inhibition of several *MMPs* (Lijnen et al., 2002). In both of these

cases, adipogenesis was halted due to impaired ECM degradation, which was dependent upon the affected metalloproteases. In their absence, the ECM became dense and fibrotic and hindered the hypertrophic expansion of adipocytes. *Adamts14* plays a role in the deposition, as opposed to degradation, of ECM components (i.e. maturation of type I collagen fibres). Regardless, the effects of the absence of *Adamts14* could inhibit adipogenesis in a similar way, as AT ECM rigidity (too much or too little) is extremely important during adipogenesis (Chun et al., 2006).

Alternatively, key-gene expression quantification on the frozen tissues might also help explain the *Adamts14*$^{-/-}$ phenotype. For example, checking WAT or BAT for the levels of expression of markers of AT beiging (i.e. *UCP1*, *-3*), would serve as an indicator of the morphological and functional shift towards canonical BAT thermogenesis (Kim et al., 2016). This phenomenon usually occurs in response to either signals from the CNS or exposure to cold temperatures. A recent study showed that exposing mice to cold triggered a change in their microbiome and also browning of WAT, increased EE and loss of AT in the short-term, with gut remodelling in the long term to increase gut size and thus nutrient absorption capacity (Chevalier et al., 2015). This animal model bears some resemblance to the

phenotype of the *Adamts14*$^{-/-}$ mice, apart from having the opposite gut length phenotype.

On the matter of rodent gut remodelling and metabolic flexibility in response to altered environments, the following model is applicable to most experimental situations. Gut flexibility arises to compensate for higher energy demands, that may be caused by pregnancy, cold exposure or low-quality diet (reviewed in (Naya, 2007)). This compensation happens through increased food intake, which in turn triggers an increase in intestinal folds and length; firstly to accommodate for the added food mass, but also to facilitate a larger nutrient absorptive area (Martinez del Rio et al., 1994). In fact, very few stimuli decrease rodent gut length, especially in the context of hyperphagia and obesity. The Lep$^{ob/ob}$ mouse for example, exhibits increased intestinal length as a response to overeating (Morton and Hanson, 1984), while HFD consumption can increase intestinal fold length in WT mice (De Wit et al., 2008).

Another study showed that mice stripped of their intestinal microbiota, i.e. germ-free (GF) mice, showed a preference to foods with lower lipid concentrations, while consuming higher volumes of them, leading to an increased total caloric intake. This was accompanied by changes in their intestinal

morphology, including a decrease in intestinal folds, which indicates impaired nutrient absorption (Duca et al., 2012). Even though not tested, this would have most likely been accompanied by a decrease in gut length. Previously, GF mice were shown to be resistant to weight gain under HFD conditions (Rabot et al., 2010), but not if exposed to a low-fat diet (Fleissner et al., 2010). They were also shown to have more "sweet" receptors and elevated expression of markers of increased sucrose uptake (Swartz et al., 2012). This phenotype is also similar to the *Adamts14*$^{-/-}$ mouse; increased caloric intake on a HFD without weight gain, indications of decreased nutrient absorption in the gut, due to decreased length, and potentially the preferential catabolism of sugars, as indicated by the higher RER I observed. This could be corroborated by checking the faecal lipid content in HFD-exposed *Adamts14*$^{-/-}$ mice.

Another study linked high-fat feeding to gastrointestinal inflammation in mice lacking *CD300a*, a leukocyte membrane-anchored glycoprotein, involved in innate immune responses, and further showed that these mice had shortened intestinal length (Tanaka et al., 2014). Thus, there is an interesting possibility that there might be an immune phenotype driving the decreased weight gain and gut length and increased EE in the *Adamts14*$^{-/-}$ mice. A general proinflammatory state

could be corroborated by checking the circulating levels of inflammatory markers (*IL1b*, *IL6*, *TNFa*, and so on). This would also be concordant with the immune activation phenotype that Dupont et al. observed in the *Adamts2*$^{-/-}$*Adamts14*$^{-/-}$ double-null animals (Dupont et al., 2018).

Altogether, this indicates that there is a lot more to be learned from the *Adamts14*$^{-/-}$ mouse. There are many theoretical possibilities that could help explain the observed phenotype, which can be tested as outlined above. However, the only clear conclusion is that this computational and experimental pipeline, designed to take GWAS signals, fine-map them to causal genes and identify their functional contributions towards body composition, has delivered interesting biology.

5.2.3 From GWAS discovery to functional characterisation

Since the first large scale GWAS in 2007 (WTCCC, 2007), the number of SNP associations to quantitative traits has sky-rocketed, from a handful to several thousands. The number of functional genomics and experimental studies of these GWAS loci is unmatched, and is the main bottleneck in elucidating the role of these signals in health and disease (Figure 5.1). The field of obesity research arguably saw one of the archetypal attempts to do so, when Claussnitzer and colleagues began

to dissect the biological basis behind the *FTO* locus association, and broke the dogma that the closest gene is most often causal, using a vast array of epigenetics, comparative genomics and gene-editing approaches (Claussnitzer et al., 2015).

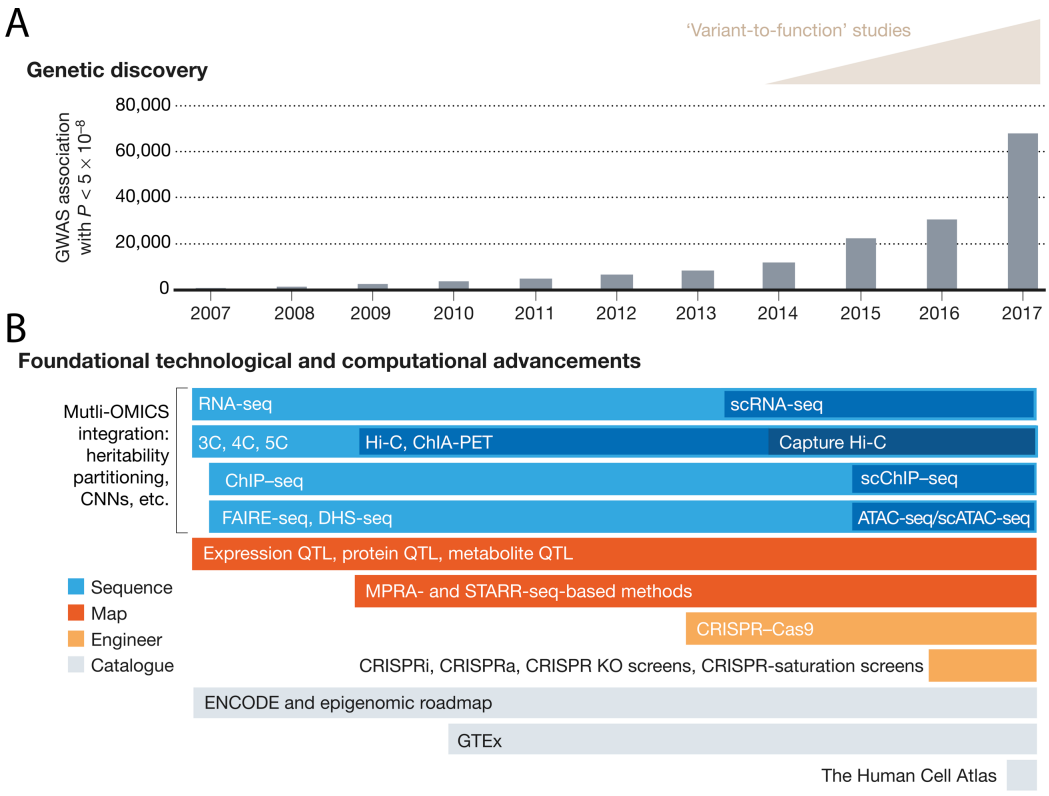


Figure 5.1 | Integration of genetic and functional breakthroughs, to aid biological understanding.

(A) Ever-increasing number of GWS association versus functional studies of said associations. (B) Recent technological advances that can facilitate more "variant-to-function" studies, including some discussed within this work. *Modified from (Claussnitzer et al., 2020).*

Other notable attempts to translate GWAS findings into biological targets have included a study which showed that a non-coding variant that associated

with myocardial infarction incidence and circulating levels of low-density lipoprotein (LDL) acted by altering the sequence of an enhancer (Musunuru et al., 2010). This alteration caused a hepatocyte-specific decrease in the expression of *SORT1* and mouse models that both over- and under-expressed this gene and exhibited altered LDL secretion patterns, validating its functional importance to lipid metabolism in mice and humans. Another study of lipid metabolism, did an exome-wide association study of circulating lipid levels and identified 242 significantly associated loci (Liu et al., 2017). In mouse models of two of these, they observed changes that phenocopy their human data. Another recent study tested the effects of knocking-out 16 replicated GWAS candidate genes for insulin resistance, in a human preadipocyte cell line (Chen et al., 2020). In doing so, they found that 12 out of 16 tested KOs affected insulin sensitisation in this cell line.

With technological advances aiding this effort in profound ways (Figure 5.1B), another recent study developed a method to integrate GWAS and mouse gene expression datasets in order to identify and prioritise conserved genes (Li et al., 2020). They did so to identify regulators of lipid metabolism, discovered 25 novel gene candidates and modelled the effects of one of them on murine metabolism, phenocopying the GWAS data and thus providing proof of principle

for their method. Following a similar approach, another study integrated human SAT transcriptomes to validate previously reported GWAS candidate loci, pinpointed additional gene candidates at them and demonstrated concordant metabolic traits in mice and humans based on the expression of their identified genes (Civelek et al., 2017).

Any such translational attempt comes with many risks and concerns. Gene candidate follow-ups hinge on identifying tissues of disease relevance, which is not always straightforward. Selecting the right model can also be complicated; whether a KO or an overexpression model or a tissue-specific mutation model (and so on) would prove most relevant for the phenotype and gene in question is hard to predict if little is known about either. As seen in the examples above, often multiple tools and models need to be attempted prior to identifying the most suitable one. In the context of this thesis, work completed on identifying a causal gene at the *PLA2G6* locus and characterising the *Adamts14*$^{-/-}$ mouse, has begun to interrogate the functional relevance of these two association signals towards body composition.

However, much more work is needed on both loci. For *PLA2G6*, we first need to verify the long-range interactions we observed in the 5C analyses. Ideally,

by modelling the specific effect of the GWAS variants in a human cell line and identifying putative effects on genome regulation and particularly the regulation of *Pick1, Dmc1* and *Sox10,* while selecting tissues that would provide the most useful adiposity insights.

For *ADAMTS14,* the null mice exhibit a fascinating phenotype, which could shed light on our human genetics findings. However, we have yet to explain the molecular basis behind their reduced weight gain and increased EE. As mentioned above, closer examination of AT from *Adamts14*$^{-/-}$ mice might show changes concordant with impaired adipogenesis or increased beiging, while expression levels of inflammatory markers and excretion levels of lipid particles would corroborate the immune activation and preferential carbohydrate utilisation theories, accordingly.

Another significant finding of this work has been the multitude of ECM remodelling genes that were significantly associated with various facets of body composition (Figure 5.2). In the iDXA GWAS, we observed GWS associations between body composition and 17 loci which include genes belonging to the *ADAM(TS)* and *MMP* gene families. Even though more work is needed to establish gene causality at these loci, given the strength of the associations and the evidence

for *Adamts14*, these GWAS candidate genes would provide a great opportunity for proof of concept work on the role of ECM remodelling genes in human adiposity.

Figure 5.2 | ECM remodellers & human adiposity.

Other members of the *MMP*, *ADAM* and *ADAMTS* gene families that significantly associated with the iDXA phenotypes. Contributions shown towards the five phenotypes with which these loci were most strongly associated and expressed as $-\log_{10}(p)$.

5.3 Concluding remark

To summarise, within this work I developed a way to efficiently impute adiposity phenotypes and used them within a GWAS framework to discover novel candidate loci. To identify causal genes at these loci I used gene expression datasets and chromatin conformation techniques. I then used a mouse model to corroborate the role of *Adamts14* in body composition. Mice with null mutations showed decreased body weight gain as well as hyperphagia and increased energy expenditure, implicating this gene in energy homeostasis. Studies, such as this, that integrate GWAS signals with functional genomic datasets and targeted alterations in model systems will rapidly increase our understanding of the genetic underpinnings of common disease.

References

ANSELME, I., LACLEF, C., LANAUD, M., RÜTHER, U. & SCHNEIDER-MAUNOURY, S. 2007. Defects in brain patterning and head morphogenesis in the mouse mutant Fused toes. *Developmental biology,* 304**,** 208-220.

ASTLE, W. J., ELDING, H., JIANG, T., ALLEN, D., RUKLISA, D., MANN, A. L., MEAD, D., BOUMAN, H., RIVEROS-MCKAY, F. & KOSTADIMA, M. A. 2016. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell,* 167**,** 1415-1429. e19.

BALL, S. D. & SWAN, P. D. 2003. ACCURACY OF ESTIMATING INTRA-ABDOMINAL FAT IN OBESE WOMEN. *Journal of Exercise Physiology Online,* 6.

BARNESS, L. A., OPITZ, J. M. & GILBERT-BARNESS, E. 2007. Obesity: genetic, molecular, and environmental aspects. *American Journal of Medical Genetics Part A,* 143**,** 3016-3034.

BEKHOUCHE, M., LEDUC, C., DUPONT, L., JANSSEN, L., DELOLME, F., VADON-LE GOFF, S., SMARGIASSO, N., BAIWIR, D., MAZZUCCHELLI, G. & ZANELLA-CLEON, I. 2016. Determination of the substrate repertoire of ADAMTS2, 3, and 14 significantly broadens their functions and identifies extracellular matrix organization and TGF-β signaling as primary targets. *The FASEB Journal,* 30**,** 1741-1756.

BENABDALLAH, N. S. & BICKMORE, W. A. Regulatory domains and their mechanisms. Cold Spring Harbor symposia on quantitative biology, 2015. Cold Spring Harbor Laboratory Press, 45-51.

BENNASAR-VENY, M., LOPEZ-GONZALEZ, A. A., TAULER, P., CESPEDES, M. L., VICENTE-HERRERO, T., YAÑEZ, A., TOMAS-SALVA, M. & AGUILO, A. 2013. Body adiposity index and cardiovascular health risk factors in Caucasians: a comparison with the body mass index and others. *PLoS One,* 8**,** e63999.

BERGMAN, R. N., STEFANOVSKI, D., BUCHANAN, T. A., SUMNER, A. E., REYNOLDS, J. C., SEBRING, N. G., XIANG, A. H. & WATANABE, R. M. 2011. A better index of body adiposity. *Obesity,* 19**,** 1083-1089.

BILLON, N., IANNARELLI, P., MONTEIRO, M. C., GLAVIEUX-PARDANAUD, C., RICHARDSON, W. D., KESSARIS, N., DANI, C. & DUPIN, E. 2007. The generation of adipocytes by the neural crest. *Development,* 134**,** 2283-2292.

BIRSOY, K., BERRY, R., WANG, T., CEYHAN, O., TAVAZOIE, S., FRIEDMAN, J. M. & RODEHEFFER, M. S. 2011. Analysis of gene networks in white adipose tissue development reveals a role for ETS2 in adipogenesis. *Development,* 138**,** 4709-4719.

BOISSEL, S., REISH, O., PROULX, K., KAWAGOE-TAKAKI, H., SEDGWICK, B., YEO, G. S., MEYRE, D., GOLZIO, C., MOLINARI, F. & KADHOM, N. 2009. Loss-of-function mutation in the dioxygenase-encoding FTO gene causes severe growth retardation and multiple malformations. *The American Journal of Human Genetics,* 85**,** 106-111.

BORATYN, G. M., SCHÄFFER, A. A., AGARWALA, R., ALTSCHUL, S. F., LIPMAN, D. J. & MADDEN, T. L. 2012. Domain enhanced lookup time accelerated BLAST. *Biology direct,* 7**,** 12.

BOURNE, R. R. & COLLABORATORS, G. R. F. 2016. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *The Lancet,* 388**,** 1659-1724.

BOYLE, E. A., LI, Y. I. & PRITCHARD, J. K. 2017. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell,* 169**,** 1177-1186.

BUNIELLO, A., MACARTHUR, J. A. L., CEREZO, M., HARRIS, L. W., HAYHURST, J., MALANGONE, C., MCMAHON, A., MORALES, J., MOUNTJOY, E. & SOLLIS, E. 2018. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic acids research,* 47**,** D1005-D1012.

BYCROFT, C., FREEMAN, C., PETKOVA, D., BAND, G., ELLIOTT, L. T., SHARP, K., MOTYER, A., VUKCEVIC, D., DELANEAU, O. & O'CONNELL, J. 2018. The UK Biobank resource with deep phenotyping and genomic data. *Nature,* 562**,** 203-209.

CARBONE, F., LA ROCCA, C. & MATARESE, G. 2012. Immunological functions of leptin and adiponectin. *Biochimie,* 94**,** 2082-2088.

CENSIN, J. C., PETERS, S. A., BOVIJN, J., FERREIRA, T., PULIT, S. L., MÄGI, R., MAHAJAN, A., HOLMES, M. V. & LINDGREN, C. M. 2019. Causal relationships between obesity and the leading causes of death in women and men. *PLoS genetics,* 15.

CHAMBERS, J. C., ZHANG, W., SEHMI, J., LI, X., WASS, M. N., VAN DER HARST, P., HOLM, H., SANNA, S., KAVOUSI, M. & BAUMEISTER, S. E. 2011. Genome-wide association study identifies loci influencing concentrations of liver enzymes in plasma. *Nature genetics,* 43**,** 1131.

CHEN, Y., CUNNINGHAM, F., RIOS, D., MCLAREN, W. M., SMITH, J., PRITCHARD, B., SPUDICH, G. M., BRENT, S., KULESHA, E. & MARIN-GARCIA, P. 2010. Ensembl variation resources. *BMC genomics,* 11**,** 293.

CHEN, Z., YU, H., SHI, X., WARREN, C. R., LOTTA, L. A., FRIESEN, M., MEISSNER, T. B., LANGENBERG, C., WABITSCH, M. & WAREHAM, N. 2020. Functional screening of candidate causal genes for insulin resistance in human preadipocytes and adipocytes. *Circulation Research,* 126**,** 330-346.

CHEVALIER, C., STOJANOVIĆ, O., COLIN, D. J., SUAREZ-ZAMORANO, N., TARALLO, V., VEYRAT-DUREBEX, C., RIGO, D., FABBIANO, S., STEVANOVIĆ, A. & HAGEMANN, S. 2015. Gut microbiota orchestrates energy homeostasis during cold. *Cell,* 163**,** 1360-1374.

CHU, A. Y., DENG, X., FISHER, V. A., DRONG, A., ZHANG, Y., FEITOSA, M. F., LIU, C.-T., WEEKS, O., CHOH, A. C. & DUAN, Q. 2017. Multiethnic genome-wide meta-analysis of ectopic fat depots identifies loci associated with adipocyte development and differentiation. *Nature genetics,* 49**,** 125.

CHUN, T.-H., HOTARY, K. B., SABEH, F., SALTIEL, A. R., ALLEN, E. D. & WEISS, S. J. 2006. A pericellular collagenase directs the 3-dimensional development of white adipose tissue. *Cell,* 125**,** 577-591.

CHUN, T.-H., INOUE, M., MORISAKI, H., YAMANAKA, I., MIYAMOTO, Y., OKAMURA, T., SATO-KUSUBATA, K. & WEISS, S. J. 2010. Genetic link between obesity and MMP14-dependent adipogenic collagen turnover. *Diabetes,* 59**,** 2484-2494.

CHUSYD, D. E., WANG, D., HUFFMAN, D. M. & NAGY, T. R. 2016. Relationships between rodent white adipose fat pads and human white adipose fat depots. *Frontiers in nutrition,* 3**,** 10.

CINTI, S., MITCHELL, G., BARBATELLI, G., MURANO, I., CERESI, E., FALOIA, E., WANG, S., FORTIER, M., GREENBERG, A. S. & OBIN, M. S. 2005. Adipocyte death defines macrophage localization and function in adipose tissue of obese mice and humans. *Journal of lipid research,* 46**,** 2347-2355.

CIVELEK, M., WU, Y., PAN, C., RAULERSON, C. K., KO, A., HE, A., TILFORD, C., SALEEM, N. K., STANČÁKOVÁ, A. & SCOTT, L. J. 2017. Genetic regulation of adipose gene expression and cardio-metabolic traits. *The American Journal of Human Genetics,* 100**,** 428-443.

CLAUSSNITZER, M., CHO, J. H., COLLINS, R., COX, N. J., DERMITZAKIS, E. T., HURLES, M. E., KATHIRESAN, S., KENNY, E. E., LINDGREN, C. M. & MACARTHUR, D. G. 2020. A brief history of human disease genetics. *Nature,* 577**,** 179-189.

CLAUSSNITZER, M., DANKEL, S. N., KIM, K.-H., QUON, G., MEULEMAN, W., HAUGEN, C., GLUNK, V., SOUSA, I. S., BEAUDRY, J. L. & PUVIINDRAN, V. 2015. FTO obesity variant circuitry and adipocyte browning in humans. *New England Journal of Medicine,* 373**,** 895-907.

COLIGE, A., VANDENBERGHE, I., THIRY, M., LAMBERT, C. A., VAN BEEUMEN, J., LI, S.-W., PROCKOP, D. J., LAPIÈRE, C. M. & NUSGENS, B. V. 2002. Cloning and characterization of ADAMTS-14, a novel ADAMTS displaying high homology with ADAMTS-2 and ADAMTS-3. *Journal of Biological Chemistry,* 277**,** 5756-5766.

COLLABORATORS, G. O. 2017. Health effects of overweight and obesity in 195 countries over 25 years. *New England Journal of Medicine,* 377**,** 13-27.

CONE, R. D. 2005. Anatomy and regulation of the central melanocortin system. *Nature neuroscience,* 8**,** 571-578.

COX, K. B., HAMM, D. A., MILLINGTON, D. S., MATERN, D., VOCKLEY, J., RINALDO, P., PINKERT, C. A., RHEAD, W. J., LINDSEY, J. R. & WOOD, P. A. 2001. Gestational, pathologic and biochemical differences between very long-chain acyl-CoA dehydrogenase deficiency and long-chain acyl-CoA dehydrogenase deficiency in the mouse. *Human molecular genetics,* 10**,** 2069-2077.

CREYGHTON, M. P., CHENG, A. W., WELSTEAD, G. G., KOOISTRA, T., CAREY, B. W., STEINE, E. J., HANNA, J., LODATO, M. A., FRAMPTON, G. M. & SHARP, P. A. 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences,* 107**,** 21931-21936.

DAHL, A., IOTCHKOVA, V., BAUD, A., JOHANSSON, Å., GYLLENSTEN, U., SORANZO, N., MOTT, R., KRANIS, A. & MARCHINI, J. 2016. A multiple-phenotype imputation method for genetic studies. *Nature genetics,* 48**,** 466.

DAY, N., OAKES, S., LUBEN, R., KHAW, K., BINGHAM, S., WELCH, A. & WAREHAM, N. 1999. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. *British journal of cancer,* 80**,** 95.

DE LUCIA ROLFE, E., LOOS, R. J., DRUET, C., STOLK, R. P., EKELUND, U., GRIFFIN, S. J., FOROUHI, N. G., WAREHAM, N. J. & ONG, K. K. 2010. Association between birth weight and visceral fat in adults–. *The American journal of clinical nutrition,* 92**,** 347-352.

DE WIT, N. J., BOSCH-VERMEULEN, H., DE GROOT, P. J., HOOIVELD, G. J., BROMHAAR, M. M. G., JANSEN, J., MÜLLER, M. & VAN DER MEER, R. 2008. The role of the small intestine in the development of dietary fat-induced obesity and insulin resistance in C57BL/6J mice. *BMC medical genomics,* 1**,** 14.

DENG, X., WANG, J., JIAO, L., UTAIPAN, T., TUMA-KELLNER, S., SCHMITZ, G., LIEBISCH, G., STREMMEL, W. & CHAMULITRAT, W. 2016. iPLA2β deficiency attenuates obesity and hepatic steatosis in ob/ob mice through hepatic fatty-acyl phospholipid remodeling. *Biochimica et Biophysica Acta (BBA)-Molecular and Cell Biology of Lipids,* 1861**,** 449-461.

DICKINSON, M. E., FLENNIKEN, A. M., JI, X., TEBOUL, L., WONG, M. D., WHITE, J. K., MEEHAN, T. F., WENINGER, W. J., WESTERBERG, H. & ADISSU, H. 2016. High-throughput discovery of novel developmental phenotypes. *Nature,* 537**,** 508.

DIGIROLAMO, M., FINE, J., TAGRA, K. & ROSSMANITH, R. 1998. Qualitative regional differences in adipose tissue growth and cellularity in male Wistar rats fed ad libitum. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* 274**,** R1460-R1467.

DIVOUX, A., TORDJMAN, J., LACASA, D., VEYRIE, N., HUGOL, D., AISSAT, A., BASDEVANT, A., GUERRE-MILLO, M., POITOU, C. & ZUCKER, J.-D. 2010. Fibrosis in human adipose tissue: composition, distribution, and link with lipid metabolism and fat mass loss. *Diabetes,* 59**,** 2817-2825.

DONG, S. S., ZHANG, Y. J., CHEN, Y. X., YAO, S., HAO, R. H., RONG, Y., NIU, H. M., CHEN, J. B., GUO, Y. & YANG, T. L. 2018. Comprehensive review and annotation of susceptibility SNPs associated with obesity-related traits. *Obesity reviews,* 19**,** 917-930.

DUCA, F. A., SWARTZ, T. D., SAKAR, Y. & COVASA, M. 2012. Increased oral detection, but decreased intestinal signaling for fats in mice lacking gut microbiota. *PLoS one,* 7.

DUFFAUT, C., ZAKAROFF-GIRARD, A., BOURLIER, V., DECAUNES, P., MAUMUS, M., CHIOTASSO, P., SENGENÈS, C., LAFONTAN, M., GALITZKY, J. & BOULOUMIÉ, A. 2009. Interplay between human adipocytes and T lymphocytes in obesity: CCL20 as an adipochemokine and T lymphocytes as lipogenic modulators. *Arteriosclerosis, thrombosis, and vascular biology,* 29**,** 1608-1614.

DUNCAN, M. J., MOTA, J., VALE, S., SANTOS, M. P. & RIBEIRO, J. C. 2013. Associations between body mass index, waist circumference and body shape

index with resting blood pressure in Portuguese adolescents. *Annals of Human Biology,* 40**,** 163-167.
DUPONT, L., EHX, G., CHANTRY, M., MONSEUR, C., LEDUC, C., JANSSEN, L., CATALDO, D., THIRY, M., JEROME, C. & THOMASSIN, J.-M. 2018. Spontaneous atopic dermatitis due to immune dysregulation in mice lacking Adamts2 and 14. *Matrix Biology,* 70**,** 140-157.
ELKS, C. E., DEN HOED, M., ZHAO, J. H., SHARP, S. J., WAREHAM, N. J., LOOS, R. J. & ONG, K. K. 2012. Variability in the heritability of body mass index: a systematic review and meta-regression. *Frontiers in endocrinology,* 3**,** 29.
EVALUATION, I. F. H. M. A. 2015. *GBD Compare* [Online]. Available: Available from http://vizhub.healthdata.org/gbd-compare. [Accessed 21/01/2020 2020].
EXIL, V. J., GARDNER, C. D., ROTTMAN, J. N., SIMS, H., BARTELDS, B., KHUCHUA, Z., SINDHAL, R., NI, G. & STRAUSS, A. W. 2006. Abnormal mitochondrial bioenergetics and heart rate dysfunction in mice lacking very-long-chain acyl-CoA dehydrogenase. *American Journal of Physiology-Heart and Circulatory Physiology,* 290**,** H1289-H1297.
FISCHER, J., KOCH, L., EMMERLING, C., VIERKOTTEN, J., PETERS, T., BRÜNING, J. C. & RÜTHER, U. 2009. Inactivation of the Fto gene protects from obesity. *Nature,* 458**,** 894.
FLEISSNER, C. K., HUEBEL, N., EL-BARY, M. M. A., LOH, G., KLAUS, S. & BLAUT, M. 2010. Absence of intestinal microbiota does not protect mice from diet-induced obesity. *British Journal of Nutrition,* 104**,** 919-929.
FOX, C. S., LIU, Y., WHITE, C. C., FEITOSA, M., SMITH, A. V., HEARD-COSTA, N., LOHMAN, K., JOHNSON, A. D., FOSTER, M. C. & GREENAWALT, D. M. 2012. Genome-wide association for abdominal subcutaneous and visceral adipose reveals a novel locus for visceral fat in women. *PLoS genetics,* 8**,** e1002695.
FRAYLING, T. M., TIMPSON, N. J., WEEDON, M. N., ZEGGINI, E., FREATHY, R. M., LINDGREN, C. M., PERRY, J. R., ELLIOTT, K. S., LANGO, H. & RAYNER, N. W. 2007. A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science,* 316**,** 889-894.
FULLER, N. 1993. Comparison of abilities of various interpretations of bio-electrical impedance to predict reference method body composition assessment. *Clinical Nutrition,* 12**,** 236-242.
FULLER, N., SAWYER, M. & ELIA, M. 1994. Comparative evaluation of body composition methods and predictions, and calculation of density and hydration fraction of fat-free mass, in obese women. *International journal of obesity,* 18**,** 503-503.
GAMAZON, E. R., WHEELER, H. E., SHAH, K. P., MOZAFFARI, S. V., AQUINO-MICHAELS, K., CARROLL, R. J., EYLER, A. E., DENNY, J. C., NICOLAE, D. L. & COX, N. J. 2015. A gene-based association method for mapping traits using reference transcriptome data. *Nature genetics,* 47**,** 1091.
GEALEKMAN, O., GUSEVA, N., HARTIGAN, C., APOTHEKER, S., GORGOGLIONE, M., GURAV, K., TRAN, K.-V., STRAUBHAAR, J., NICOLORO, S. & CZECH, M. P. 2011. Depot-specific differences and insufficient subcutaneous adipose tissue angiogenesis in human obesity. *Circulation,* 123**,** 186-194.

GESTA, S., TSENG, Y.-H. & KAHN, C. R. 2007. Developmental origin of fat: tracking obesity to its source. *Cell,* 131**,** 242-256.

GOLDSTEIN, B. J., SCALIA, R. G. & MA, X. L. 2009. Protective vascular and myocardial effects of adiponectin. *Nature clinical practice Cardiovascular medicine,* 6**,** 27-35.

GREENFIELD, J. R., SAMARAS, K., CHISHOLM, D. J. & CAMPBELL, L. V. 2002. Regional intra-subject variability in abdominal adiposity limits usefulness of computed tomography. *Obesity research,* 10**,** 260-265.

GUO, T., JOU, W., CHANTURIYA, T., PORTAS, J., GAVRILOVA, O. & MCPHERRON, A. C. 2009. Myostatin inhibition in muscle, but not adipose tissue, decreases fat mass and improves insulin sensitivity. *PloS one,* 4.

HALLER, T., KALS, M., ESKO, T., MÄGI, R. & FISCHER, K. 2015. RegScan: a GWAS tool for quick estimation of allele effects on continuous traits and their combinations. *Briefings in bioinformatics,* 16**,** 39-44.

HAN, J., LEE, J.-E., JIN, J., LIM, J. S., OH, N., KIM, K., CHANG, S.-I., SHIBUYA, M., KIM, H. & KOH, G. Y. 2011. The spatiotemporal development of adipose tissue. *Development,* 138**,** 5027-5037.

HE, Q., GAO, Z., YIN, J., ZHANG, J., YUN, Z. & YE, J. 2011. Regulation of HIF-1α activity in adipose tissue by obesity-associated factors: adipogenesis, insulin, and hypoxia. *American Journal of Physiology-Endocrinology and Metabolism,* 300**,** E877-E885.

HE, S. & CHEN, X. 2013. Could the new body shape index predict the new onset of diabetes mellitus in the Chinese population? *PLoS One,* 8**,** e50573.

HEART, N., LUNG, INSTITUTE, B., DIABETES, N. I. O., DIGESTIVE & DISEASES, K. 1998. *Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: the evidence report*, National Heart, Lung, and Blood Institute.

HEID, I. M., JACKSON, A. U., RANDALL, J. C., WINKLER, T. W., QI, L., STEINTHORSDOTTIR, V., THORLEIFSSON, G., ZILLIKENS, M. C., SPELIOTES, E. K. & MÄGI, R. 2010. Meta-analysis identifies 13 new loci associated with waist-hip ratio and reveals sexual dimorphism in the genetic basis of fat distribution. *Nature genetics,* 42**,** 949-960.

HEINTZMAN, N. D., STUART, R. K., HON, G., FU, Y., CHING, C. W., HAWKINS, R. D., BARRERA, L. O., VAN CALCAR, S., QU, C. & CHING, K. A. 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature genetics,* 39**,** 311-318.

HERRERA, B. M., KEILDSON, S. & LINDGREN, C. M. 2011. Genetics and epigenetics of obesity. *Maturitas,* 69**,** 41-49.

HEYMSFIELD, S. B., GALLAGHER, D., MAYER, L., BEETSCH, J. & PIETROBELLI, A. 2007. Scaling of human body composition to stature: new insights into body mass index. *The American journal of clinical nutrition,* 86**,** 82-91.

HILL JR, R. B., PROSPER, J., HIRSCHFIELD, J. S. & KERN JR, F. 1968. Protein starvation and the small intestine: I. The growth and morphology of the small intestine in weanling rats. *Experimental and molecular pathology,* 8**,** 66-74.

HORMOZDIARI, F., KANG, E. Y., BILOW, M., BEN-DAVID, E., VULPE, C., MCLACHLAN, S., LUSIS, A. J., HAN, B. & ESKIN, E. 2016. Imputing phenotypes for genome-wide association studies. *The American Journal of Human Genetics,* 99**,** 89-103.

IANNUZZI-SUCICH, M., PRESTWOOD, K. M. & KENNY, A. M. 2002. Prevalence of sarcopenia and predictors of skeletal muscle mass in healthy, older men and women. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences,* 57**,** M772-M777.

JANSSEN, L., DUPONT, L., BEKHOUCHE, M., NOEL, A., LEDUC, C., VOZ, M., PEERS, B., CATALDO, D., APTE, S. S. & DUBAIL, J. 2016. ADAMTS3 activity is mandatory for embryonic lymphangiogenesis and regulates placental angiogenesis. *Angiogenesis,* 19**,** 53-65.

JOSHI, P. K., PIRASTU, N., KENTISTOU, K. A., FISCHER, K., HOFER, E., SCHRAUT, K. E., CLARK, D. W., NUTILE, T., BARNES, C. L. & TIMMERS, P. R. 2017. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nature communications,* 8**,** 910.

JOSHI, P. K., PRENDERGAST, J., FRASER, R. M., HUFFMAN, J. E., VITART, V., HAYWARD, C., MCQUILLAN, R., GLODZIK, D., POLAŠEK, O. & HASTIE, N. D. 2013. Local exome sequences facilitate imputation of less common variants and increase power of genome wide association studies. *PloS one,* 8**,** e68604.

JUSTICE, A. E., KARADERI, T., HIGHLAND, H. M., YOUNG, K. L., GRAFF, M., LU, Y., TURCOT, V., AUER, P. L., FINE, R. S. & GUO, X. 2019. Protein-coding variants implicate novel genes related to lipid homeostasis contributing to body-fat distribution. *Nature genetics***,** 1.

KAJIMURA, S., SPIEGELMAN, B. M. & SEALE, P. 2015. Brown and beige fat: physiological roles beyond heat generation. *Cell metabolism,* 22**,** 546-559.

KAJIRO, M., TSUCHIYA, M., KAWABE, Y.-I., FURUMAI, R., IWASAKI, N., HAYASHI, Y., KATANO, M., NAKAJIMA, Y., GOTO, N. & WATANABE, T. 2011. The E3 ubiquitin ligase activity of Trip12 is essential for mouse embryogenesis. *PLoS One,* 6.

KANNEGANTI, T.-D. & DIXIT, V. D. 2012. Immunological complications of obesity. *Nature immunology,* 13**,** 707-712.

KARLSSON, T., RASK-ANDERSEN, M., PAN, G., HÖGLUND, J., WADELIUS, C., EK, W. E. & JOHANSSON, Å. 2019. Contribution of genetics to visceral adiposity and its relation to cardiovascular and metabolic disease. *Nature medicine,* 25**,** 1390-1395.

KARPE, F. & PINNICK, K. E. 2015. Biology of upper-body and lower-body adipose tissue [mdash] link to whole-body phenotypes. *Nature reviews Endocrinology,* 11**,** 90-100.

KAWAGUCHI, N., XU, X., TAJIMA, R., KRONQVIST, P., SUNDBERG, C., LOECHEL, F., ALBRECHTSEN, R. & WEWER, U. M. 2002. ADAM 12 protease induces adipogenesis in transgenic mice. *The American journal of pathology,* 160**,** 1895-1903.

KELLY, T. L., WILSON, K. E. & HEYMSFIELD, S. B. 2009. Dual energy X-Ray absorptiometry body composition reference values from NHANES. *PloS one,* 4**,** e7038.

KENTISTOU, K. A., WILSON, J. F., JOSHI, P. K. & MORTON, N. M. 2019. The genetic underpinnings of obesity. *Current Opinion in Physiology*.

KETTUNEN, J., TUKIAINEN, T., SARIN, A.-P., ORTEGA-ALONSO, A., TIKKANEN, E., LYYTIKÄINEN, L.-P., KANGAS, A. J., SOININEN, P., WÜRTZ, P. & SILANDER, K. 2012. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. *Nature genetics,* 44**,** 269.

KHOKHA, R., MURTHY, A. & WEISS, A. 2013. Metalloproteinases and their natural inhibitors in inflammation and immunity. *Nature Reviews Immunology,* 13**,** 649-665.

KIM, C. A., DELÉPINE, M., BOUTET, E., EL MOURABIT, H., LE LAY, S., MEIER, M., NEMANI, M., BRIDEL, E., LEITE, C. C. & BERTOLA, D. R. 2008. Association of a homozygous nonsense caveolin-1 mutation with Berardinelli-Seip congenital lipodystrophy. *The Journal of Clinical Endocrinology & Metabolism,* 93**,** 1129-1134.

KIM, J.-Y., VAN DE WALL, E., LAPLANTE, M., AZZARA, A., TRUJILLO, M. E., HOFMANN, S. M., SCHRAW, T., DURAND, J. L., LI, H. & LI, G. 2007. Obesity-associated improvements in metabolic profile through expansion of adipose tissue. *The Journal of clinical investigation,* 117**,** 2621-2637.

KIM, S., YU, N.-K. & KAANG, B.-K. 2015. CTCF as a multifunctional protein in genome regulation and gene expression. *Experimental & molecular medicine,* 47**,** e166-e166.

KIM, S.-N., JUNG, Y.-S., KWON, H.-J., SEONG, J. K., GRANNEMAN, J. G. & LEE, Y.-H. 2016. Sex differences in sympathetic innervation and browning of white adipose tissue of mice. *Biology of sex differences,* 7**,** 67.

KNITTLE, J., TIMMERS, K., GINSBERG-FELLNER, F., BROWN, R. & KATZ, D. 1979. The growth of adipose tissue in children and adolescents. Cross-sectional and longitudinal studies of adipose cell number and size. *The Journal of clinical investigation,* 63**,** 239-246.

KRAKAUER, N. Y. & KRAKAUER, J. C. 2012. A new body shape index predicts mortality hazard independently of body mass index. *PloS one,* 7**,** e39504.

KU, M., KOCHE, R. P., RHEINBAY, E., MENDENHALL, E. M., ENDOH, M., MIKKELSEN, T. S., PRESSER, A., NUSBAUM, C., XIE, X. & CHI, A. S. 2008. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS genetics,* 4.

KUMAR, S., AMBROSINI, G. & BUCHER, P. 2016. SNP2TFBS–a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic acids research,* 45**,** D139-D144.

KURISAKI, T., MASUDA, A., SUDO, K., SAKAGAMI, J., HIGASHIYAMA, S., MATSUDA, Y., NAGABUKURO, A., TSUJI, A., NABESHIMA, Y. & ASANO, M. 2003. Phenotypic analysis of Meltrin α (ADAM12)-deficient mice: involvement of Meltrin α in adipogenesis and myogenesis. *Molecular and cellular biology,* 23**,** 55-61.

KYLE, U. G., BOSAEUS, I., DE LORENZO, A. D., DEURENBERG, P., ELIA, M., GÓMEZ, J. M., HEITMANN, B. L., KENT-SMITH, L., MELCHIOR, J.-C. & PIRLICH, M. 2004. Bioelectrical impedance analysis—part II: utilization in clinical practice. *Clinical nutrition,* 23**,** 1430-1453.

LAJOIE, B. R., VAN BERKUM, N. L., SANYAL, A. & DEKKER, J. 2009. My5C: web tools for chromosome conformation capture studies. *Nature methods,* 6**,** 690-691.

LEPPER, C. & FAN, C. M. 2010. Inducible lineage tracing of Pax7-descendant cells reveals embryonic origin of adult satellite cells. *genesis,* 48**,** 424-436.

LEVITT, D. G., BECKMAN, L. M., MAGER, J. R., VALENTINE, B., SIBLEY, S. D., BECKMAN, T. R., KELLOGG, T. A., IKRAMUDDIN, S. & EARTHMAN, C. P. 2010. Comparison of DXA and water measurements of body fat following gastric bypass surgery and a physiological model of body water, fat, and muscle composition. *Journal of applied physiology,* 109**,** 786-795.

LI, G., XIE, C., LU, S., NICHOLS, R. G., TIAN, Y., LI, L., PATEL, D., MA, Y., BROCKER, C. N. & YAN, T. 2017. Intermittent fasting promotes white adipose browning and decreases obesity by shaping the gut microbiota. *Cell metabolism,* 26**,** 672-685. e4.

LI, S.-W., ARITA, M., FERTALA, A., BAO, Y., KOPEN, G. C., LÅNGSJÖ, T. K., HYTTINEN, M. M., HELMINEN, H. J. & PROCKOP, D. J. 2001. Transgenic mice with inactive alleles for procollagen N-proteinase (ADAMTS-2) develop fragile skin and male sterility. *Biochemical Journal,* 355**,** 271-278.

LI, Z., VOTAVA, J. A., ZAJAC, G. J., NGUYEN, J. N., JAIMES, F. B. L., LY, S. M., BRINKMAN, J. A., DE GIORGI, M., KAUL, S. & GREEN, C. L. 2020. Integrating Mouse and Human Genetic Data to Move beyond GWAS and Identify Causal Genes in Cholesterol Metabolism. *Cell Metabolism*.

LICHTASH, C. T., CUI, J., GUO, X., CHEN, Y.-D. I., HSUEH, W. A., ROTTER, J. I. & GOODARZI, M. O. 2013. Body adiposity index versus body mass index and other anthropometric traits as correlates of cardiometabolic risk factors. *PLoS One,* 8**,** e65954.

LIJNEN, H., MAQUOI, E., HANSEN, L., VAN HOEF, B., FREDERIX, L. & COLLEN, D. 2002. Matrix metalloproteinase inhibition impairs adipose tissue development in mice. *Arteriosclerosis, thrombosis, and vascular biology,* 22**,** 374-379.

LIU, B., GLOUDEMANS, M. J., RAO, A. S., INGELSSON, E. & MONTGOMERY, S. B. 2019. Abundant associations with gene expression complicate GWAS follow-up. *Nature genetics,* 51**,** 768-769.

LIU, D. J., PELOSO, G. M., YU, H., BUTTERWORTH, A. S., WANG, X., MAHAJAN, A., SALEHEEN, D., EMDIN, C., ALAM, D. & ALVES, A. C. 2017. Exome-wide association study of plasma lipids in> 300,000 individuals. *Nature genetics,* 49**,** 1758.

LIU, J., FOX, C. S., HICKSON, D. A., MAY, W. D., HAIRSTON, K. G., CARR, J. J. & TAYLOR, H. A. 2010. Impact of abdominal visceral and subcutaneous adipose tissue on cardiometabolic risk factors: the Jackson Heart Study. *The Journal of Clinical Endocrinology & Metabolism,* 95**,** 5419-5426.

LIU, Y.-J., LIU, X.-G., WANG, L., DINA, C., YAN, H., LIU, J.-F., LEVY, S., PAPASIAN, C. J., DREES, B. M. & HAMILTON, J. J. 2008. Genome-wide association scans

identified CTNNBL1 as a novel gene for obesity. *Human molecular genetics,* 17**,** 1803-1813.
LLOYD-JONES, L. R., HOLLOWAY, A., MCRAE, A., YANG, J., SMALL, K., ZHAO, J., ZENG, B., BAKSHI, A., METSPALU, A. & DERMITZAKIS, M. 2017. The genetic architecture of gene expression in peripheral blood. *The American Journal of Human Genetics,* 100**,** 228-237.
LOCKE, A. E., KAHALI, B., BERNDT, S. I., JUSTICE, A. E., PERS, T. H., DAY, F. R., POWELL, C., VEDANTAM, S., BUCHKOVICH, M. L. & YANG, J. 2015. Genetic studies of body mass index yield new insights for obesity biology. *Nature,* 518**,** 197.
LONSDALE, J., THOMAS, J., SALVATORE, M., PHILLIPS, R., LO, E., SHAD, S., HASZ, R., WALTERS, G., GARCIA, F. & YOUNG, N. 2013. The genotype-tissue expression (GTEx) project. *Nature genetics,* 45**,** 580.
LOOS, R. J. 2018. The genetics of adiposity. *Current opinion in genetics & development,* 50**,** 86-95.
LÓPEZ, A. A., CESPEDES, M. L., VICENTE, T., TOMAS, M., BENNASAR-VENY, M., TAULER, P. & AGUILO, A. 2012. Body adiposity index utilization in a Spanish Mediterranean population: comparison with the body mass index. *PloS one,* 7**,** e35281.
LOTTA, L. A., GULATI, P., DAY, F. R., PAYNE, F., ONGEN, H., VAN DE BUNT, M., GAULTON, K. J., EICHER, J. D., SHARP, S. J. & LUAN, J. A. 2017. Integrative genomic analysis implicates limited peripheral adipose storage capacity in the pathogenesis of human insulin resistance. *Nature genetics,* 49**,** 17.
LU, Y., DAY, F. R., GUSTAFSSON, S., BUCHKOVICH, M. L., NA, J., BATAILLE, V., COUSMINER, D. L., DASTANI, Z., DRONG, A. W. & ESKO, T. 2016. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nature communications,* 7**,** 10495.
MAEDA, N., SHIMOMURA, I., KISHIDA, K., NISHIZAWA, H., MATSUDA, M., NAGARETANI, H., FURUYAMA, N., KONDO, H., TAKAHASHI, M. & ARITA, Y. 2002. Diet-induced insulin resistance in mice lacking adiponectin/ACRP30. *Nature medicine,* 8**,** 731-737.
MAESSEN, M. F., EIJSVOGELS, T. M., VERHEGGEN, R. J., HOPMAN, M. T., VERBEEK, A. L. & DE VEGT, F. 2014. Entering a new era of body indices: the feasibility of a body shape index and body roundness index to identify cardiovascular health status. *PLoS One,* 9**,** e107212.
MANOLIO, T. A., COLLINS, F. S., COX, N. J., GOLDSTEIN, D. B., HINDORFF, L. A., HUNTER, D. J., MCCARTHY, M. I., RAMOS, E. M., CARDON, L. R. & CHAKRAVARTI, A. 2009. Finding the missing heritability of complex diseases. *Nature,* 461**,** 747-753.
MARTINEZ DEL RIO, C., CORK, S. J. & KARASOV, W. H. 1994. Modelling gut function: an introduction.
MCQUILLAN, R., LEUTENEGGER, A.-L., ABDEL-RAHMAN, R., FRANKLIN, C. S., PERICIC, M., BARAC-LAUC, L., SMOLEJ-NARANCIC, N., JANICIJEVIC, B., POLASEK, O. & TENESA, A. 2008. Runs of homozygosity in European populations. *The American Journal of Human Genetics,* 83**,** 359-372.

MONTAGUE, C. T., FAROOQI, I. S., WHITEHEAD, J. P. & SOOS, M. A. 1997. Congenital leptin deficiency is associated with severe early-onset obesity in humans. *Nature,* 387**,** 903.

MONTEIRO, C. A., MOURA, E. C., CONDE, W. L. & POPKIN, B. M. 2004. Socioeconomic status and obesity in adult populations of developing countries: a review. *Bulletin of the World Health Organization,* 82**,** 940-946.

MORTENSEN, E. L. K., WANG, T., MALTE, H., RAUBENHEIMER, D. & MAYNTZ, D. 2010. Maternal preconceptional nutrition leads to variable fat deposition and gut dimensions of adult offspring mice (C57BL/6JBom). *International journal of obesity,* 34**,** 1618-1624.

MORTON, A. & HANSON, P. 1984. Monosaccharide transport by the small intestine of lean and genetically obese (ob/ob) mice. *Quarterly Journal of Experimental Physiology: Translation and Integration,* 69**,** 117-126.

MOURTZAKIS, M., PRADO, C. M., LIEFFERS, J. R., REIMAN, T., MCCARGAR, L. J. & BARACOS, V. E. 2008. A practical and precise approach to quantification of body composition in cancer patients using computed tomography images acquired during routine care. *Applied Physiology, Nutrition, and Metabolism,* 33**,** 997-1006.

MUSUNURU, K., STRONG, A., FRANK-KAMENETSKY, M., LEE, N. E., AHFELDT, T., SACHS, K. V., LI, X., LI, H., KUPERWASSER, N. & RUDA, V. M. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature,* 466**,** 714-719.

MYTTON, O. T., OGILVIE, D., GRIFFIN, S., BRAGE, S., WAREHAM, N. & PANTER, J. 2018. Associations of active commuting with body fat and visceral adipose tissue: A cross-sectional population based study in the UK. *Preventive medicine,* 106**,** 86-93.

NAGAI, M., KOMIYA, H., MORI, Y., OHTA, T., KASAHARA, Y. & IKEDA, Y. 2008. Development of a new method for estimating visceral fat area with multi-frequency bioelectrical impedance. *The Tohoku journal of experimental medicine,* 214**,** 105-112.

NAGASE, H., VISSE, R. & MURPHY, G. 2006. Structure and function of matrix metalloproteinases and TIMPs. *Cardiovascular research,* 69**,** 562-573.

NAKAJIMA, I., MUROYA, S., TANABE, R. I. & CHIKUNI, K. 2002. Extracellular matrix development during differentiation into adipocytes with a unique increase in type V and VI collagen. *Biology of the Cell,* 94**,** 197-203.

NAYA, D. 2007. Phenotypic plasticity in laboratory mice and rats: a meta-analysis of current ideas on gut size flexibility.

NCEP 2001. Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *Jama,* 285**,** 2486.

NEVILLE, M. J., WITTEMANS, L. B., PINNICK, K. E., TODORČEVIĆ, M., KAKSONEN, R., PIETILÄINEN, K. H., LUAN, J. A., SCOTT, R. A., WAREHAM, N. J. & LANGENBERG, C. 2019. Regional fat depot masses are influenced by protein-coding gene variants. *BioRxiv***,** 526434.

NEWTON-CHEH, C., EIJGELSHEIM, M., RICE, K. M., DE BAKKER, P. I., YIN, X., ESTRADA, K., BIS, J. C., MARCIANTE, K., RIVADENEIRA, F. & NOSEWORTHY, P. A. 2009. Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nature genetics,* 41**,** 399.

NISHIMURA, S., MANABE, I., NAGASAKI, M., HOSOYA, Y., YAMASHITA, H., FUJITA, H., OHSUGI, M., TOBE, K., KADOWAKI, T. & NAGAI, R. 2007. Adipogenesis in obesity requires close interplay between differentiating adipocytes, stromal cells, and blood vessels. *Diabetes,* 56**,** 1517-1526.

ONG, K. K., ELKS, C. E., LI, S., ZHAO, J. H., LUAN, J. A., ANDERSEN, L. B., BINGHAM, S. A., BRAGE, S., SMITH, G. D. & EKELUND, U. 2009. Genetic variation in LIN28B is associated with the timing of puberty. *Nature genetics,* 41**,** 729.

OTTAVIANI, E., MALAGOLI, D. & FRANCESCHI, C. 2011. The evolution of the adipose tissue: a neglected enigma. *General and comparative endocrinology,* 174**,** 1-4.

PARK, D. & YOON, M. 2012. Compound K, a novel ginsenoside metabolite, inhibits adipocyte differentiation in 3T3-L1 cells: involvement of angiogenesis and MMPs. *Biochemical and biophysical research communications,* 422**,** 263-267.

PASARICA, M., SEREDA, O. R., REDMAN, L. M., ALBARADO, D. C., HYMEL, D. T., ROAN, L. E., ROOD, J. C., BURK, D. H. & SMITH, S. R. 2009. Reduced adipose tissue oxygenation in human obesity: evidence for rarefaction, macrophage chemotaxis, and inflammation without an angiogenic response. *Diabetes,* 58**,** 718-725.

PASCO, J. A., HOLLOWAY, K. L., DOBBINS, A. G., KOTOWICZ, M. A., WILLIAMS, L. J. & BRENNAN, S. L. 2014. Body mass index and measures of body fat for defining obesity and underweight: a cross-sectional, population-based study. *BMC obesity,* 1**,** 9.

PATTARO, C., TEUMER, A., GORSKI, M., CHU, A. Y., LI, M., MIJATOVIC, V., GARNAAS, M., TIN, A., SORICE, R. & LI, Y. 2016. Genetic associations at 53 loci highlight cell types and biological pathways relevant for kidney function. *Nature communications,* 7**,** 10023.

PELLEGRINELLI, V., CAROBBIO, S. & VIDAL-PUIG, A. 2016. Adipose tissue plasticity: how fat depots respond differently to pathophysiological cues. *Diabetologia,* 59**,** 1075-1088.

PENDÁS, A. M., FOLGUERAS, A. R., LLANO, E., CATERINA, J., FRERARD, F., RODRÍGUEZ, F., ASTUDILLO, A., NOËL, A., BIRKEDAL-HANSEN, H. & LÓPEZ-OTÍN, C. 2004. Diet-induced obesity and reduced skin cancer susceptibility in matrix metalloproteinase 19-deficient mice. *Molecular and cellular biology,* 24**,** 5304-5313.

POISSONNET, C. M., BURDI, A. R. & GARN, S. M. 1984. The chronology of adipose tissue appearance and distribution in the human fetus.

PORCU, E., RÜEGER, S., LEPIK, K., SANTONI, F. A., REYMOND, A. & KUTALIK, Z. 2019. Mendelian Randomization integrating GWAS and eQTL data reveals genetic determinants of complex and clinical traits. *bioRxiv***,** 377267.

PULIT, S. L., KARADERI, T. & LINDGREN, C. M. 2017. Sexual dimorphisms in genetic loci linked to body fat distribution. *Bioscience reports,* 37**,** BSR20160184.

PULIT, S. L., STONEMAN, C., MORRIS, A. P., WOOD, A. R., GLASTONBURY, C. A., TYRRELL, J., YENGO, L., FERREIRA, T., MAROULI, E. & JI, Y. 2018. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Human molecular genetics,* 28**,** 166-174.

RABOT, S., MEMBREZ, M., BRUNEAU, A., GÉRARD, P., HARACH, T., MOSER, M., RAYMOND, F., MANSOURIAN, R. & CHOU, C. J. 2010. Germ-free C57BL/6J mice are resistant to high-fat-diet-induced insulin resistance and have altered cholesterol metabolism. *The FASEB Journal,* 24**,** 4948-4959.

RASK-ANDERSEN, M., KARLSSON, T., EK, W. E. & JOHANSSON, Å. 2019. Genome-wide association study of body fat distribution identifies adiposity loci and sex-specific genetic effects. *Nature communications,* 10**,** 339.

RODRIGUEZ-LOPEZ, J., POMBO-SUAREZ, M., LOUGHLIN, J., TSEZOU, A., BLANCO, F., MEULENBELT, I., SLAGBOOM, P., VALDES, A., SPECTOR, T. & GOMEZ-REINO, J. 2009. Association of a nsSNP in ADAMTS14 to some osteoarthritis phenotypes. *Osteoarthritis and cartilage,* 17**,** 321-327.

RONCO, A. L. & DE STÉFANI, E. 2012. Anthropometry. *Nutritional Epidemiology of Breast Cancer.* Springer.

ROSENBLOOM, K. R., SLOAN, C. A., MALLADI, V. S., DRESZER, T. R., LEARNED, K., KIRKUP, V. M., WONG, M. C., MADDREN, M., FANG, R. & HEITNER, S. G. 2012. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic acids research,* 41**,** D56-D63.

SASSI, F., DEVAUX, M., CECCHINI, M. & RUSTICELLI, E. 2009. The obesity epidemic: analysis of past and projected future trends in selected OECD countries.

SAUNDERS, C. L., CHIODINI, B. D., SHAM, P., LEWIS, C. M., ABKEVICH, V., ADEYEMO, A. A., DE ANDRADE, M., ARYA, R., BERENSON, G. S. & BLANGERO, J. 2007. Meta-analysis of genome-wide linkage studies in BMI and obesity. *Obesity,* 15**,** 2263-2275.

SAVAGE, D. B., TAN, G. D., ACERINI, C. L., JEBB, S. A., AGOSTINI, M., GURNELL, M., WILLIAMS, R. L., UMPLEBY, A. M., THOMAS, E. L. & BELL, J. D. 2003. Human metabolic syndrome resulting from dominant-negative mutations in the nuclear receptor peroxisome proliferator-activated receptor-γ. *Diabetes,* 52**,** 910-917.

SCAFOGLIERI, A., CLARYS, J. P., BAUER, J. M., VERLAAN, S., VAN MALDEREN, L., VANTIEGHEM, S., CEDERHOLM, T., SIEBER, C. C., METS, T. & BAUTMANS, I. 2017. Predicting appendicular lean and fat mass with bioelectrical impedance analysis in older adults with physical function decline–The PROVIDE study. *Clinical Nutrition,* 36**,** 869-875.

SCHOFIELD, J. P., COX, T. M., CASKEY, C. T. & WAKAMIYA, M. 1999. Mice deficient in the urea-cycle enzyme, carbamoyl phosphate synthetase i, die during the early neonatal period from hyperammonemia. *Hepatology,* 29**,** 181-185.

SCOTT, R. A., FALL, T., PASKO, D., BARKER, A., SHARP, S. J., ARRIOLA, L., BALKAU, B., BARRICARTE, A., BARROSO, I. & BOEING, H. 2014. Common genetic variants highlight the role of insulin resistance and body fat distribution in type 2 diabetes, independent of obesity. *Diabetes,* 63**,** 4378-4387.

SEALE, P., BJORK, B., YANG, W., KAJIMURA, S., CHIN, S., KUANG, S., SCIME, A., DEVARAKONDA, S., CONROE, H. M. & ERDJUMENT-BROMAGE, H. 2008. PRDM16 controls a brown fat/skeletal muscle switch. *Nature,* 454**,** 961-967.

SHE, X., ROHL, C. A., CASTLE, J. C., KULKARNI, A. V., JOHNSON, J. M. & CHEN, R. 2009. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC genomics,* 10**,** 269.

SHINDO, T., KURIHARA, H., KUNO, K., YOKOYAMA, H., WADA, T., KURIHARA, Y., IMAI, T., WANG, Y., OGATA, M. & NISHIMATSU, H. 2000. ADAMTS-1: a metalloproteinase-disintegrin essential for normal growth, fertility, and organ morphology and function. *The Journal of clinical investigation,* 105**,** 1345-1352.

SHUNGIN, D., WINKLER, T. W., CROTEAU-CHONKA, D. C., FERREIRA, T., LOCKE, A. E., MÄGI, R., STRAWBRIDGE, R. J., PERS, T. H., FISCHER, K. & JUSTICE, A. E. 2015. New genetic loci link adipose and insulin biology to body fat distribution. *Nature,* 518**,** 187.

SHUSTER, A., PATLAS, M., PINTHUS, J. & MOURTZAKIS, M. 2012. The clinical importance of visceral adiposity: a critical review of methods for visceral adipose tissue analysis. *The British journal of radiology,* 85**,** 1-10.

SILVER, H. J., WELCH, E. B., AVISON, M. J. & NISWENDER, K. D. 2010. Imaging body composition in obesity and weight loss: challenges and opportunities. *Diabetes, metabolic syndrome and obesity: targets and therapy,* 3**,** 337.

SMEMO, S., TENA, J. J., KIM, K.-H., GAMAZON, E. R., SAKABE, N. J., GÓMEZ-MARÍN, C., ANEAS, I., CREDIDIO, F. L., SOBREIRA, D. R. & WASSERMAN, N. F. 2014. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature,* 507**,** 371.

SORANZO, N., RIVADENEIRA, F., CHINAPPEN-HORSLEY, U., MALKINA, I., RICHARDS, J. B., HAMMOND, N., STOLK, L., NICA, A., INOUYE, M. & HOFMAN, A. 2009. Meta-analysis of genome-wide scans for human adult stature identifies novel Loci and associations with measures of skeletal frame size. *PLoS genetics,* 5**,** e1000445.

SPENCER, M., YAO-BORENGASSER, A., UNAL, R., RASOULI, N., GURLEY, C. M., ZHU, B., PETERSON, C. A. & KERN, P. A. 2010. Adipose tissue macrophages in insulin-resistant subjects are associated with collagen VI and fibrosis and demonstrate alternative activation. *American Journal of Physiology-Endocrinology and Metabolism,* 299**,** E1016-E1027.

SPICUGLIA, S. & VANHILLE, L. 2012. Chromatin signatures of active enhancers. *Nucleus,* 3**,** 126-131.

SPRACKLEN, C. N., CHEN, P., KIM, Y. J., WANG, X., CAI, H., LI, S., LONG, J., WU, Y., WANG, Y. X. & TAKEUCHI, F. 2017. Association analyses of East Asian individuals and trans-ancestry analyses with European individuals reveal new loci associated with cholesterol and triglyceride levels. *Human molecular genetics,* 26**,** 1770-1784.

STEEL, J. H., WHITE, R. & PARKER, M. G. 2005. Role of the RIP140 corepressor in ovulation and adipose biology. *Journal of endocrinology,* 185**,** 1-9.

STELZER, G., ROSEN, N., PLASCHKES, I., ZIMMERMAN, S., TWIK, M., FISHILEVICH, S., STEIN, T. I., NUDEL, R., LIEDER, I. & MAZOR, Y. 2016. The GeneCards suite: from gene data mining to disease genome sequence analyses. *Current protocols in bioinformatics,* 54**,** 1.30. 1-1.30. 33.

STEVENS, J., CAI, J., PAMUK, E. R., WILLIAMSON, D. F., THUN, M. J. & WOOD, J. L. 1998. The effect of age on the association between body-mass index and mortality. *New England Journal of Medicine,* 338**,** 1-7.

STRATIGOPOULOS, G., CARLI, J. F. M., O'DAY, D. R., WANG, L., LEDUC, C. A., LANZANO, P., CHUNG, W. K., ROSENBAUM, M., EGLI, D. & DOHERTY, D. A. 2014. Hypomorphism for RPGRIP1L, a ciliary gene vicinal to the FTO locus, causes increased adiposity in mice. *Cell metabolism,* 19**,** 767-779.

STRATIGOPOULOS, G., PADILLA, S. L., LEDUC, C. A., WATSON, E., HATTERSLEY, A. T., MCCARTHY, M. I., ZELTSER, L. M., CHUNG, W. K. & LEIBEL, R. L. 2008. Regulation of Fto/Ftm gene expression in mice and humans. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology,* 294**,** R1185-R1196.

SUDLOW, C., GALLACHER, J., ALLEN, N., BERAL, V., BURTON, P., DANESH, J., DOWNEY, P., ELLIOTT, P., GREEN, J. & LANDRAY, M. 2015. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine,* 12**,** e1001779.

SWARTZ, T. D., DUCA, F., DE WOUTERS, T., SAKAR, Y. & COVASA, M. 2012. Up-regulation of intestinal type 1 taste receptor 3 and sodium glucose luminal transporter-1 expression and increased sucrose intake in mice lacking gut microbiota. *British Journal of Nutrition,* 107**,** 621-630.

SWINBURN, B. A., SACKS, G., HALL, K. D., MCPHERSON, K., FINEGOOD, D. T., MOODIE, M. L. & GORTMAKER, S. L. 2011. The global obesity pandemic: shaped by global drivers and local environments. *The Lancet,* 378**,** 804-814.

TANAKA, T., TAHARA-HANAOKA, S., NABEKURA, T., IKEDA, K., JIANG, S., TSUTSUMI, S., INAGAKI, T., MAGOORI, K., HIGURASHI, T. & TAKAHASHI, H. 2014. PPARβ/δ activation of CD300a controls intestinal immunity. *Scientific reports,* 4**,** 5412.

TCHKONIA, T., THOMOU, T., ZHU, Y., KARAGIANNIDES, I., POTHOULAKIS, C., JENSEN, M. D. & KIRKLAND, J. L. 2013. Mechanisms and metabolic implications of regional differences among fat depots. *Cell metabolism,* 17**,** 644-656.

TCHOUKALOVA, Y. D., VOTRUBA, S. B., TCHKONIA, T., GIORGADZE, N., KIRKLAND, J. L. & JENSEN, M. D. 2010. Regional differences in cellular mechanisms of adipose tissue gain with overfeeding. *Proceedings of the National Academy of Sciences,* 107**,** 18226-18231.

TESLOVICH, T. M., MUSUNURU, K., SMITH, A. V., EDMONDSON, A. C., STYLIANOU, I. M., KOSEKI, M., PIRRUCCELLO, J. P., RIPATTI, S., CHASMAN, D. I. & WILLER, C. J. 2010. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature,* 466**,** 707.

THOMAS, D. M., BREDLAU, C., BOSY-WESTPHAL, A., MUELLER, M., SHEN, W., GALLAGHER, D., MAEDA, Y., MCDOUGALL, A., PETERSON, C. M. & RAVUSSIN, E. 2013. Relationships between body roundness with body fat and visceral

adipose tissue emerging from a new geometrical model. *Obesity,* 21**,** 2264-2271.
THOMAS, E. L., SAEED, N., HAJNAL, J. V., BRYNES, A., GOLDSTONE, A. P., FROST, G. & BELL, J. D. 1998. Magnetic resonance imaging of total body fat. *Journal of Applied Physiology,* 85**,** 1778-1785.
THOMOU, T., TCHKONIA, T. & KIRKLAND, J. L. 2010. Cellular and molecular basis of functional differences among fat depots. *Adipose tissue in health and disease***,** 21-47.
TIMMONS, J. A., WENNMALM, K., LARSSON, O., WALDEN, T. B., LASSMANN, T., PETROVIC, N., HAMILTON, D. L., GIMENO, R. E., WAHLESTEDT, C. & BAAR, K. 2007. Myogenic gene expression signature establishes that brown and white adipocytes originate from distinct cell lineages. *Proceedings of the National Academy of Sciences,* 104**,** 4401-4406.
TURCOT, V., LU, Y., HIGHLAND, H. M., SCHURMANN, C., JUSTICE, A. E., FINE, R. S., BRADFIELD, J. P., ESKO, T., GIRI, A. & GRAFF, M. 2018. Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nature genetics,* 50**,** 26.
TURER, A. T., KHERA, A., AYERS, C., TURER, C. B., GRUNDY, S. M., VEGA, G. L. & SCHERER, P. E. 2011. Adipose tissue mass and location affect circulating adiponectin levels. *Diabetologia,* 54**,** 2515.
UHLÉN, M., FAGERBERG, L., HALLSTRÖM, B. M., LINDSKOG, C., OKSVOLD, P., MARDINOGLU, A., SIVERTSSON, Å., KAMPF, C., SJÖSTEDT, E. & ASPLUND, A. 2015. Tissue-based map of the human proteome. *Science,* 347**,** 1260419.
VAISSE, C., CLEMENT, K., CABROLL, S., LAHLOU, N., BASDEVANT, A., BOUGNERES, P., LEBOUC, Y., GUY-GRAND, B. & FROGUEL, P. 1998. An Homozygous Mutation in the Human Leptin Receptor Leads to Morbid Obesity and Pituitary Dysfunction, but not Diabetes. *Diabetes,* 47**,** 14A.
VAN HUL, M., BAUTERS, D. & LIJNEN, R. H. 2013. Differential effects of a gelatinase inhibitor on adipocyte differentiation and adipose tissue development. *Clinical and Experimental Pharmacology and Physiology,* 40**,** 689-697.
VAN HUL, M. & LIJNEN, H. 2008. A functional role of gelatinase A in the development of nutritionally induced obesity in mice. *Journal of Thrombosis and Haemostasis,* 6**,** 1198-1206.
VAN MEURS, J. B., PARE, G., SCHWARTZ, S. M., HAZRA, A., TANAKA, T., VERMEULEN, S. H., COTLARCIUC, I., YUAN, X., MÄLARSTIG, A. & BANDINELLI, S. 2013. Common genetic loci influencing plasma homocysteine concentrations and their effect on risk of coronary artery disease–. *The American journal of clinical nutrition,* 98**,** 668-676.
VAZQUEZ, G., DUVAL, S., JACOBS JR, D. R. & SILVENTOINEN, K. 2007. Comparison of body mass index, waist circumference, and waist/hip ratio in predicting incident diabetes: a meta-analysis. *Epidemiologic reviews,* 29**,** 115-128.
VÕSA, U., CLARINGBOULD, A., WESTRA, H.-J., BONDER, M. J., DEELEN, P., ZENG, B., KIRSTEN, H., SAHA, A., KREUZHUBER, R. & KASELA, S. 2018. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *BioRxiv***,** 447367.

WANG, Q. A., TAO, C., GUPTA, R. K. & SCHERER, P. E. 2013. Tracking adipogenesis during white adipose tissue development, expansion and regeneration. *Nature medicine,* 19**,** 1338.

WEISBERG, S. P., MCCANN, D., DESAI, M., ROSENBAUM, M., LEIBEL, R. L. & FERRANTE, A. W. 2003. Obesity is associated with macrophage accumulation in adipose tissue. *The Journal of clinical investigation,* 112**,** 1796-1808.

WELLS, J. C. 2007. Sexual dimorphism of body composition. *Best practice & research Clinical endocrinology & metabolism,* 21**,** 415-430.

WESTRA, H.-J., PETERS, M. J., ESKO, T., YAGHOOTKAR, H., SCHURMANN, C., KETTUNEN, J., CHRISTIANSEN, M. W., FAIRFAX, B. P., SCHRAMM, K. & POWELL, J. E. 2013. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature genetics,* 45**,** 1238-1243.

WHO 2000. *Obesity: preventing and managing the global epidemic*, World Health Organization.

WHO 2011. Waist circumference and waist-hip ratio: Report of a WHO expert consultation, Geneva, 8-11 December 2008.

WILLER, C. J., SCHMIDT, E. M., SENGUPTA, S., PELOSO, G. M., GUSTAFSSON, S., KANONI, S., GANNA, A., CHEN, J., BUCHKOVICH, M. L. & MORA, S. 2013. Discovery and refinement of loci associated with lipid levels. *Nature genetics,* 45**,** 1274.

WINKLER, T. W., GÜNTHER, F., HÖLLERER, S., ZIMMERMANN, M., LOOS, R. J. F., KUTALIK, Z. & HEID, I. M. 2018. A joint view on genetic variants for adiposity differentiates subtypes with distinct metabolic implications. *Nature Communications,* 9**,** 1946.

WINKLER, T. W., JUSTICE, A. E., GRAFF, M., BARATA, L., FEITOSA, M. F., CHU, S., CZAJKOWSKI, J., ESKO, T., FALL, T. & KILPELÄINEN, T. O. 2015. The influence of age and sex on genetic associations with adult body size and shape: a large-scale genome-wide interaction study. *PLoS genetics,* 11**,** e1005378.

WOODS, S. C. & D'ALESSIO, D. A. 2008. Central control of body weight and appetite. *The Journal of Clinical Endocrinology & Metabolism,* 93**,** s37-s50.

WTCCC 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature,* 447**,** 661.

YAGHOOTKAR, H., SCOTT, R. A., WHITE, C. C., ZHANG, W., SPELIOTES, E., MUNROE, P. B., EHRET, G. B., BIS, J. C., FOX, C. S. & WALKER, M. 2014. Genetic evidence for a normal-weight "metabolically obese" phenotype linking insulin resistance, hypertension, coronary artery disease, and type 2 diabetes. *Diabetes,* 63**,** 4369-4377.

YAZDI, F. T., CLEE, S. M. & MEYRE, D. 2015. Obesity genetics in mouse and human: back and forth, and back again. *PeerJ,* 3**,** e856.

YENGO, L., SIDORENKO, J., KEMPER, K. E., ZHENG, Z., WOOD, A. R., WEEDON, M. N., FRAYLING, T. M., HIRSCHHORN, J., YANG, J. & VISSCHER, P. M. 2018. Meta-analysis of genome-wide association studies for height and body mass index in~ 700000 individuals of European ancestry. *Human molecular genetics,* 27**,** 3641-3649.

YUSUF, S., HAWKEN, S., OUNPUU, S., BAUTISTA, L., FRANZOSI, M. G., COMMERFORD, P., LANG, C. C., RUMBOLDT, Z., ONEN, C. L. & LISHENG, L. 2005. Obesity and the risk of myocardial infarction in 27 000 participants from 52 countries: a case-control study. *The Lancet,* 366**,** 1640-1649.

ZARAGOSI, L.-E., WDZIEKONSKI, B., VILLAGEOIS, P., KEOPHIPHATH, M., MAUMUS, M., TCHKONIA, T., BOURLIER, V., MOHSEN-KANSON, T., LADOUX, A. & ELABD, C. 2010. Activin a plays a critical role in proliferation and differentiation of human adipose progenitors. *Diabetes,* 59**,** 2513-2521.

ZENTNER, G. E., TESAR, P. J. & SCACHERI, P. C. 2011. Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome research,* 21**,** 1273-1283.

ZHANG, L., ZHONG, S., LI, Y., JI, G., SUNDARAM, M. & YAO, Z. 2013. Global inactivation of the Pla2g6 gene in mice does not cause dyslipidemia under chow or high-fat diet conditions. *Journal of cancer prevention,* 18**,** 235.

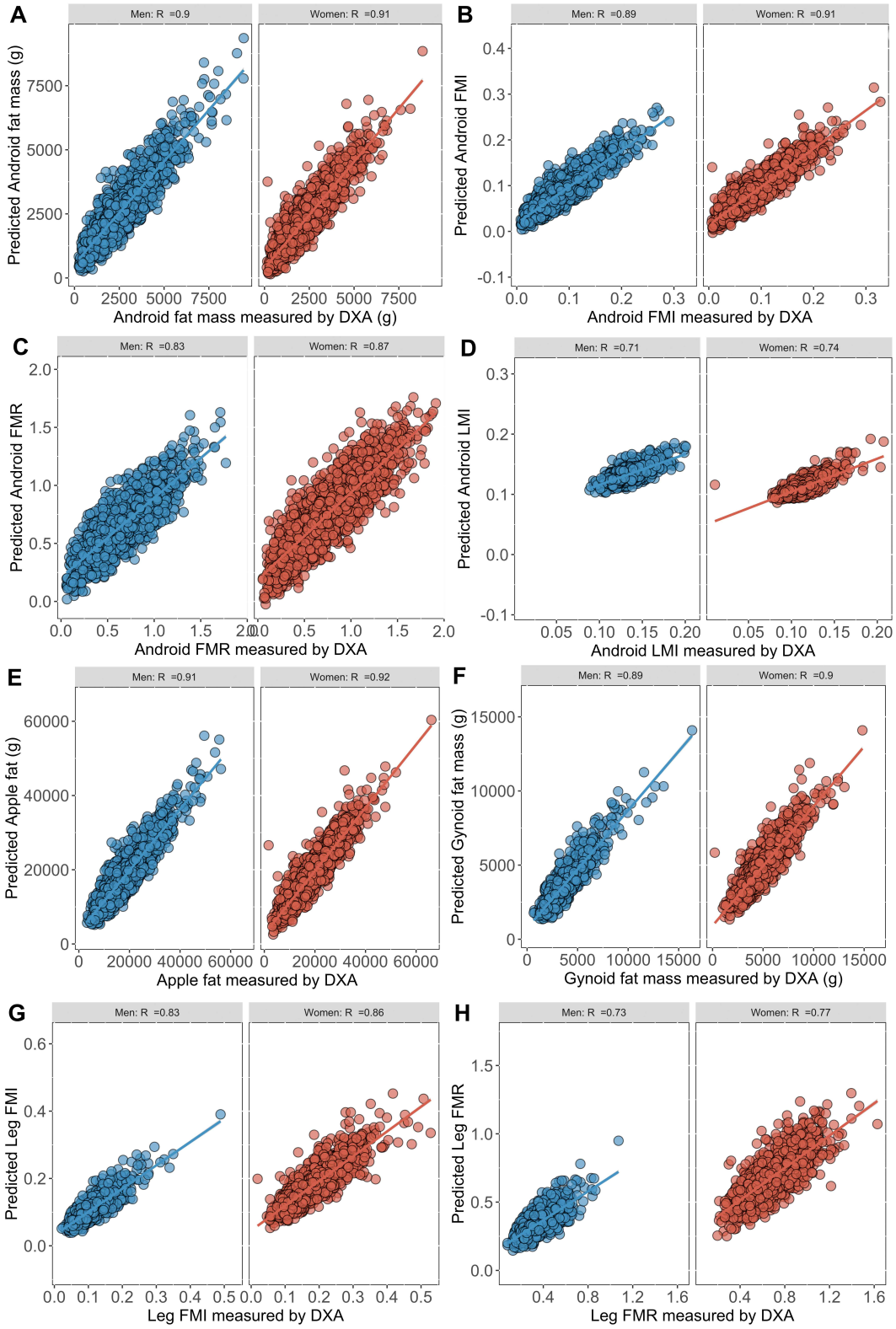
ZHANG, X., BAILEY, S. D. & LUPIEN, M. 2014. Laying a solid foundation for Manhattan–'setting the functional basis for the post-GWAS era'. *Trends in Genetics,* 30**,** 140-149.

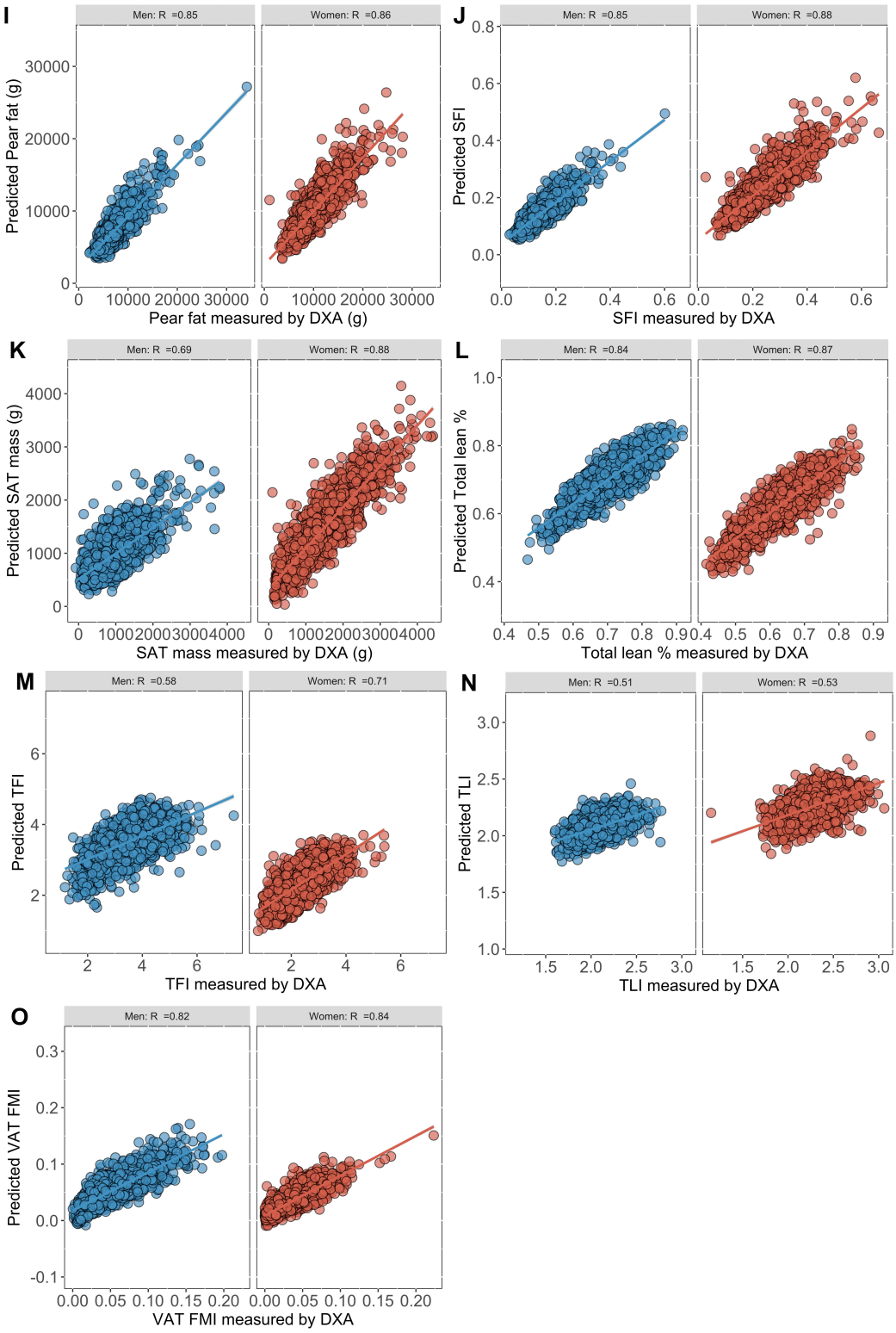
ZHANG, Y., PROENCA, R., MAFFEI, M., BARONE, M., LEOPOLD, L. & FRIEDMAN, J. M. 1994. Positional cloning of the mouse obese gene and its human homologue. *Nature,* 372**,** 425-432.

ZHENG, J., RICHARDSON, T., MILLARD, L., HEMANI, G., RAISTRICK, C., VILHJALMSSON, B., HAYCOCK, P. & GAUNT, T. 2017. PhenoSpD: an integrated toolkit for phenotypic correlation estimation and multiple testing correction using GWAS summary statistics. *bioRxiv*.

ZHU, Z., ZHANG, F., HU, H., BAKSHI, A., ROBINSON, M. R., POWELL, J. E., MONTGOMERY, G. W., GODDARD, M. E., WRAY, N. R. & VISSCHER, P. M. 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nature genetics,* 48**,** 481.

Figure S1 | Reasonably strong phenotypic correlations between DXA measurements and iDXA estimates.





Sex-separated predictions of the DXA phenotypes using BIA and anthropometric phenotypes in the 5,000 participants, with DXA measured in UKB. Correlations depicted using Pearson's R.

Figure S2 | DXA and iDXA correlations with metabolic health in women and men.

Phenotypic correlations between obesity disease sequelae and DXA phenotypes in women and men (A & B) and iDXA (C & D).

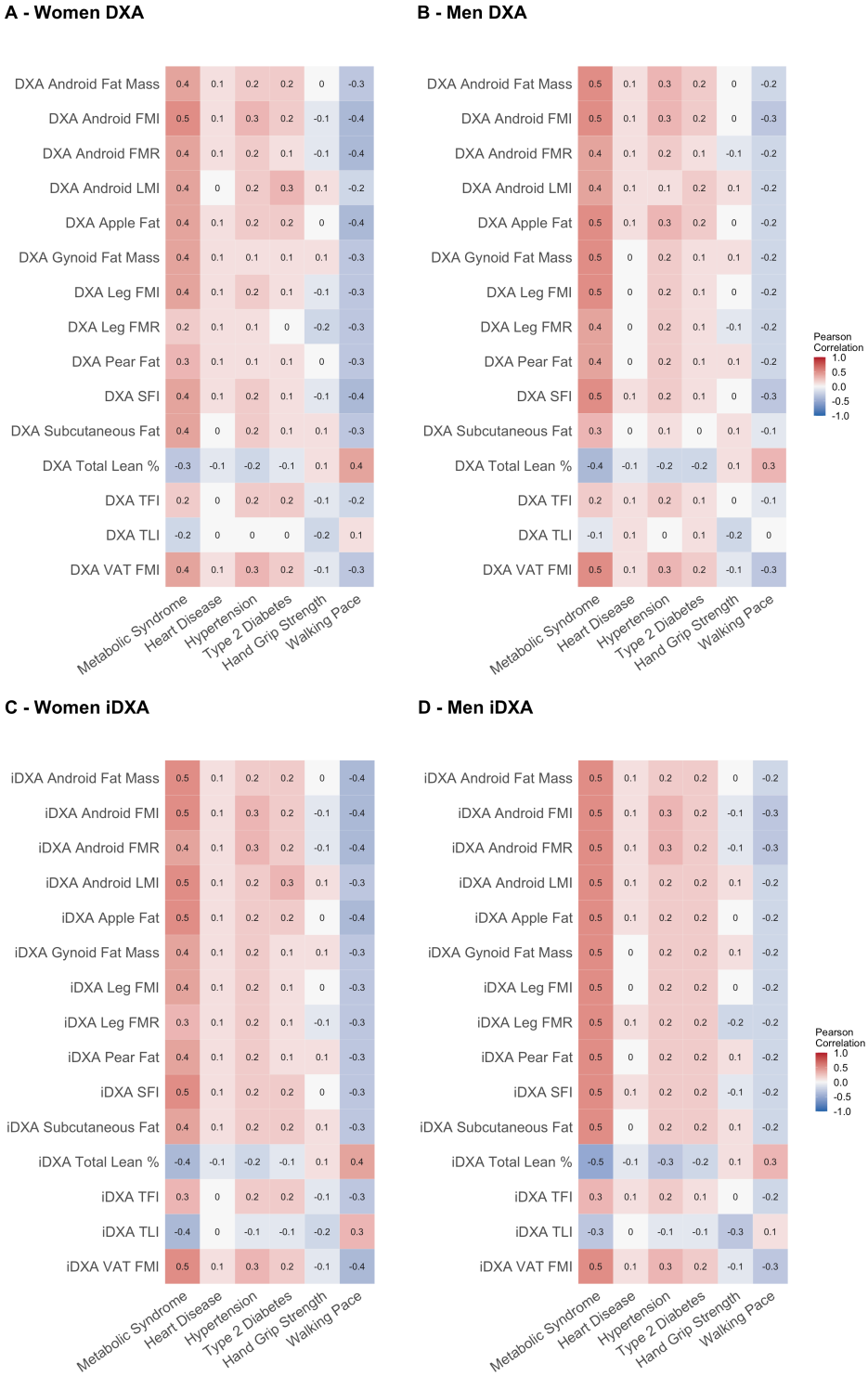
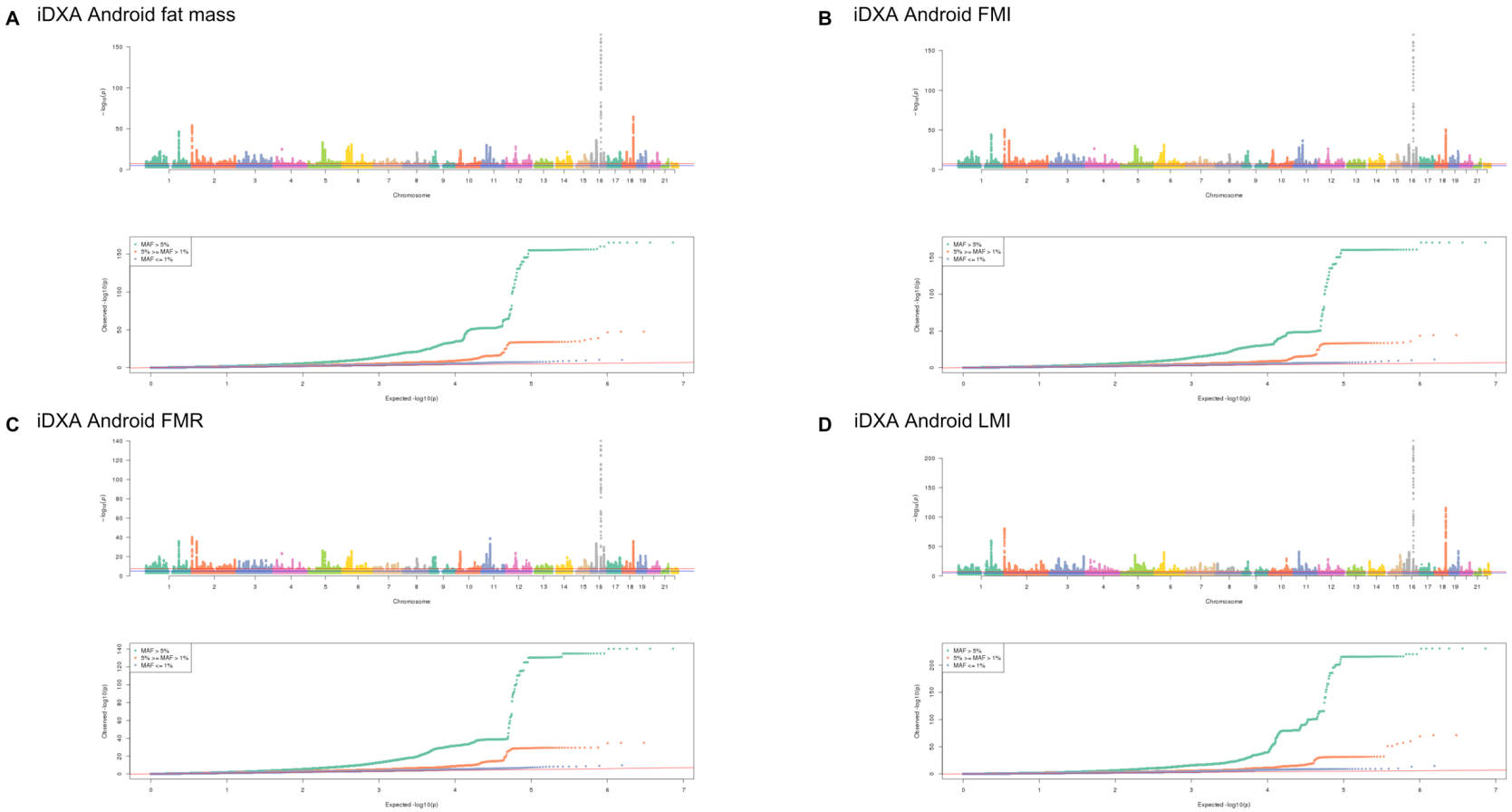
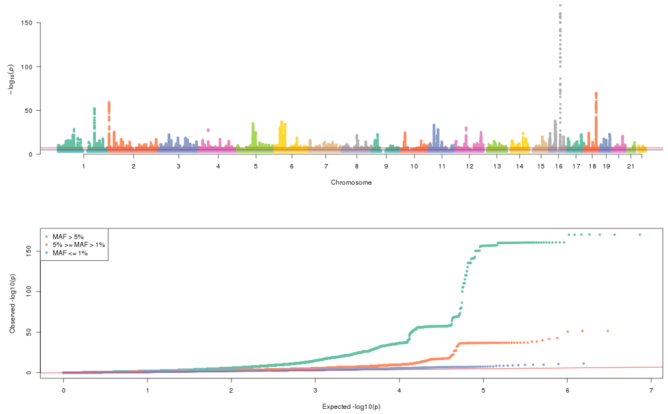


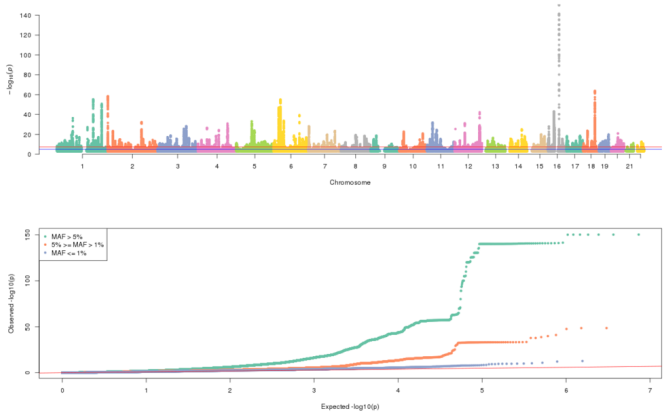
Figure S3 | Manhattan and QQ plots for the iDXA GWAS.



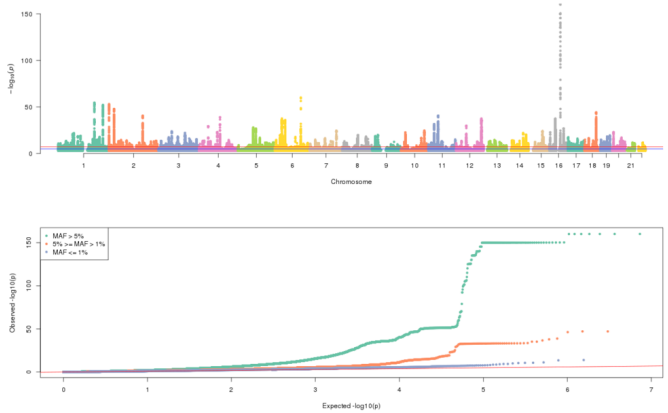
E iDXA Apple fat



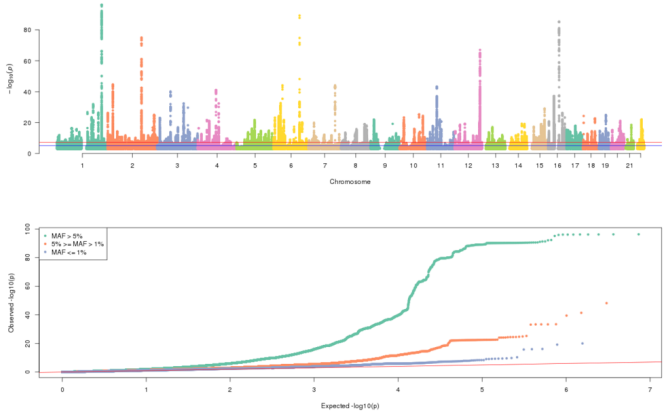
F iDXA Gynoid fat mass



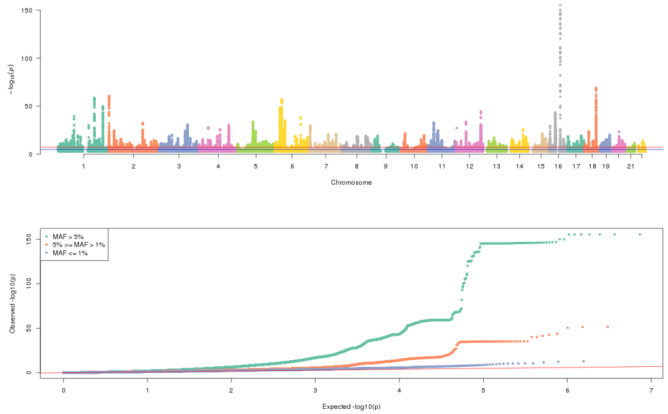
G iDXA Leg FMI



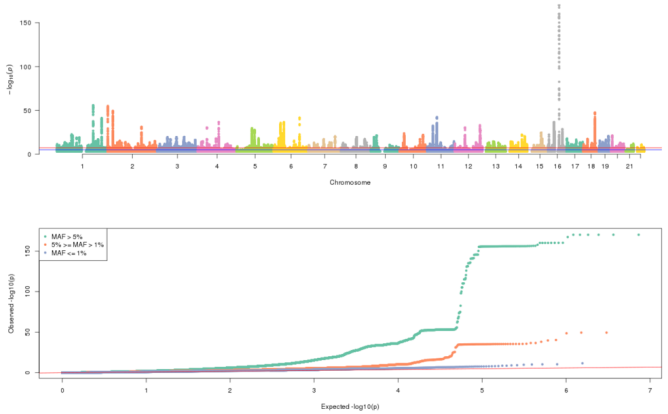
H iDXA Leg FMR



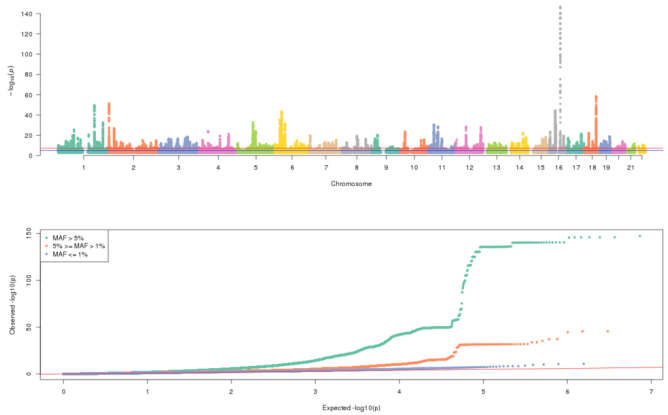
I iDXA Pear fat



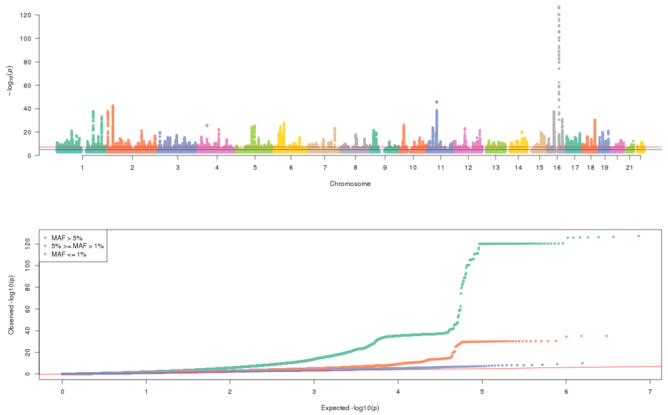
J iDXA SFI



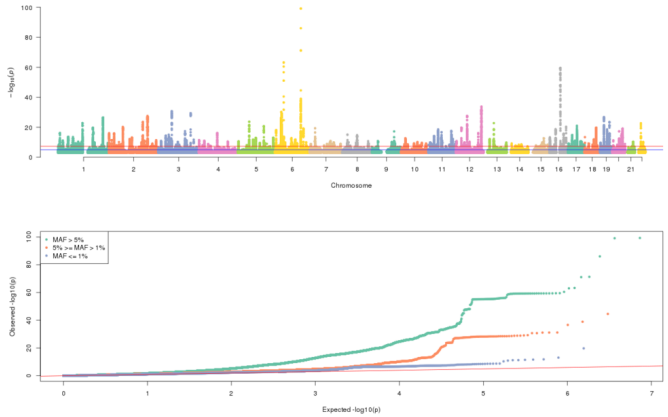
K iDXA Subcutaneous fat



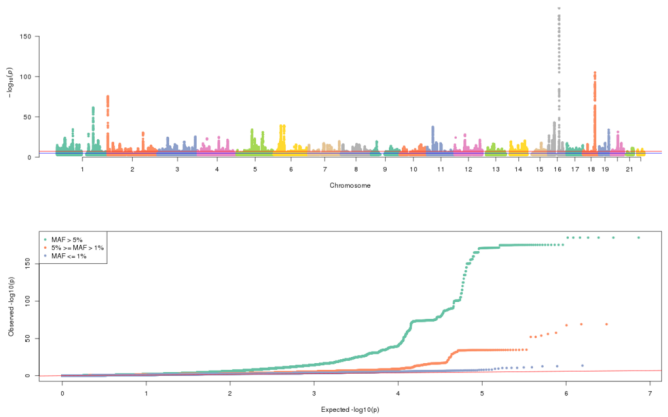
L iDXA Total lean %



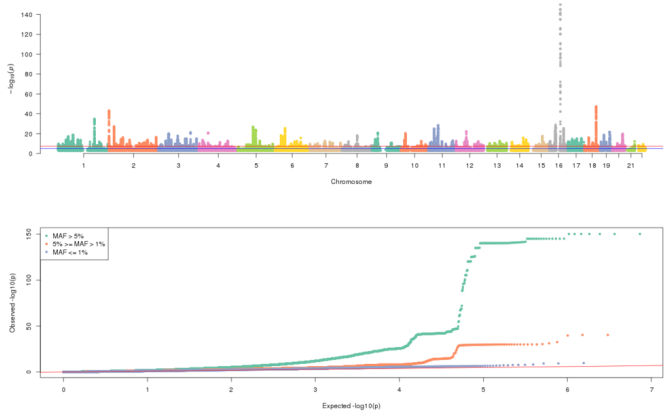
M iDXA TFI



N iDXA TLI



O iDXA VAT FMI



The 15 prediction models above, were used to calculate the iDXA in the rest of UKB. GWAS were then conducted on them after excluding all non-white-British and all related individuals, in a remaining cohort of approximately 390,000 individuals. Genome-wide Manhattan and QQ plots depict the observed SNP-trait associations and the distributions of the observed versus expected GWAS p-values.

Table S1 | Linear model coefficients for covariates used in the iDXA GWAS models.

Prior to running GWAS, the iDXA phenotypes were corrected for several demographic and study design confounders. Below is a table of the effect magnitudes of these covariates on iDXA android fat mass. This information can be accessed via https://doi.org/10.7488/ds/2918 for all other queried iDXA phenotypes. Significance codes: '***' < 0.001, '**'< 0.01, '*' < 0.05.

Coefficients	*Estimate*	*SE*	*t value*	*Pr(>|t|)*	
(Intercept)	3045.0	122.0	25.0	<2E-16	***
age	3.0	0.2	12.1	<2E-16	***
sex	-550.2	3.8	-144.4	<2E-16	***
ass_cBirmingham	551.5	37.9	14.6	<2E-16	***
ass_cBristol	481.6	39.3	12.3	<2E-16	***
ass_cBury	542.5	59.1	9.2	<2E-16	***
ass_cCardiff	714.4	48.1	14.9	<2E-16	***
ass_cCroydon	439.8	18.2	24.2	<2E-16	***
ass_cEdinburgh	469.0	112.0	4.2	2.83E-05	***
ass_cGlasgow	518.7	116.1	4.5	7.96E-06	***
ass_cHounslow	334.1	18.3	18.3	<2E-16	***
ass_cLeeds	599.3	58.9	10.2	<2E-16	***
ass_cLiverpool	596.6	61.0	9.8	<2E-16	***
ass_cManchester	537.0	56.7	9.5	<2E-16	***
ass_cMiddlesborough	485.9	75.1	6.5	9.90E-11	***
ass_cNewcastle	613.6	84.2	7.3	3.09E-13	***
ass_cNottingham	618.1	40.5	15.3	<2E-16	***
ass_cOxford	540.0	25.8	20.9	<2E-16	***
ass_cReading	583.6	21.0	27.8	<2E-16	***
ass_cSheffield	651.8	49.9	13.1	<2E-16	***
ass_cStockport	604.1	82.3	7.3	2.17E-13	***
ass_cStoke	609.0	49.0	12.4	<2E-16	***
ass_cSwansea	612.0	64.5	9.5	<2E-16	***
ass_cWrexham	605.3	76.2	7.9	1.89E-15	***
town_di	129.8	2.2	60.1	<2E-16	***
edu	-18.4	0.4	-48.8	<2E-16	***
northing	0.0	0.0	-0.7	0.4805	
easting	0.0	0.0	1.3	0.201	
genotyping.arrayUKBL	163.4	27.1	6.0	1.68E-09	***

batchBatch_b002	26.7	27.5	1.0	0.3309
batchBatch_b003	-11.4	27.5	-0.4	0.6772
batchBatch_b004	-5.7	27.5	-0.2	0.8359
batchBatch_b005	27.3	27.5	1.0	0.3214
batchBatch_b006	30.2	27.3	1.1	0.2686
batchBatch_b007	30.9	27.4	1.1	0.2598
batchBatch_b008	6.7	27.3	0.2	0.8069
batchBatch_b009	19.3	27.4	0.7	0.4813
batchBatch_b010	-7.9	27.5	-0.3	0.7727
batchBatch_b011	-0.3	27.3	0.0	0.9923
batchBatch_b012	15.8	27.3	0.6	0.5632
batchBatch_b013	8.7	27.4	0.3	0.7492
batchBatch_b014	-16.2	27.3	-0.6	0.5539
batchBatch_b015	-40.6	27.4	-1.5	0.1382
batchBatch_b016	31.0	27.6	1.1	0.2625
batchBatch_b017	26.1	27.6	0.9	0.3444
batchBatch_b018	2.8	27.5	0.1	0.9178
batchBatch_b019	18.5	27.6	0.7	0.5012
batchBatch_b020	-5.6	27.6	-0.2	0.8381
batchBatch_b021	24.9	27.7	0.9	0.3682
batchBatch_b022	27.1	27.4	1.0	0.3231
batchBatch_b023	-3.8	27.5	-0.1	0.8887
batchBatch_b024	37.3	27.4	1.4	0.174
batchBatch_b025	27.8	27.4	1.0	0.3102
batchBatch_b026	-4.7	27.4	-0.2	0.8625
batchBatch_b027	35.3	27.5	1.3	0.1982
batchBatch_b028	28.7	27.4	1.0	0.2958
batchBatch_b029	-21.2	27.4	-0.8	0.4394
batchBatch_b030	3.5	27.5	0.1	0.8991
batchBatch_b031	-0.8	27.4	0.0	0.9774
batchBatch_b032	4.4	27.3	0.2	0.8722
batchBatch_b033	32.4	27.5	1.2	0.238
batchBatch_b034	23.4	27.4	0.9	0.3937
batchBatch_b035	3.5	27.5	0.1	0.8974
batchBatch_b036	35.7	27.3	1.3	0.1912
batchBatch_b037	-9.1	27.5	-0.3	0.7419
batchBatch_b038	18.4	27.5	0.7	0.5019
batchBatch_b039	27.7	27.5	1.0	0.3138

batchBatch_b040	8.4	27.3	0.3	0.7594	
batchBatch_b041	52.7	27.5	1.9	0.0554	.
batchBatch_b042	-46.9	27.5	-1.7	0.0887	.
batchBatch_b043	4.8	27.4	0.2	0.8614	
batchBatch_b044	-25.4	27.4	-0.9	0.3538	
batchBatch_b045	35.8	27.5	1.3	0.1925	
batchBatch_b046	-23.3	27.6	-0.8	0.3978	
batchBatch_b047	16.6	27.5	0.6	0.5449	
batchBatch_b048	-12.6	27.3	-0.5	0.6435	
batchBatch_b049	-34.3	27.5	-1.2	0.2122	
batchBatch_b050	18.0	27.4	0.7	0.5105	
batchBatch_b051	40.3	27.4	1.5	0.1404	
batchBatch_b052	21.3	27.7	0.8	0.4425	
batchBatch_b053	9.7	27.4	0.4	0.7245	
batchBatch_b054	2.6	27.5	0.1	0.9258	
batchBatch_b055	28.4	27.6	1.0	0.3033	
batchBatch_b056	16.7	27.4	0.6	0.5431	
batchBatch_b057	-13.3	27.6	-0.5	0.63	
batchBatch_b058	62.4	27.4	2.3	0.0229	*
batchBatch_b059	-2.5	27.4	-0.1	0.9258	
batchBatch_b060	-0.4	27.4	0.0	0.9869	
batchBatch_b061	-9.0	27.5	-0.3	0.7438	
batchBatch_b062	10.7	27.5	0.4	0.6967	
batchBatch_b063	12.8	27.4	0.5	0.6398	
batchBatch_b064	54.4	27.4	2.0	0.0473	*
batchBatch_b065	21.9	27.5	0.8	0.425	
batchBatch_b066	-5.8	27.6	-0.2	0.8322	
batchBatch_b067	19.5	27.4	0.7	0.478	
batchBatch_b068	32.5	27.5	1.2	0.2362	
batchBatch_b069	30.9	27.5	1.1	0.261	
batchBatch_b070	0.6	27.4	0.0	0.9825	
batchBatch_b071	13.6	27.5	0.5	0.6223	
batchBatch_b072	22.0	27.5	0.8	0.4226	
batchBatch_b073	27.3	27.5	1.0	0.3201	
batchBatch_b074	19.7	27.5	0.7	0.4737	
batchBatch_b075	29.3	27.5	1.1	0.2867	
batchBatch_b076	38.3	27.3	1.4	0.1612	
batchBatch_b077	28.3	27.5	1.0	0.3034	

batchBatch_b078	16.3	27.3	0.6	0.5518	
batchBatch_b079	-7.1	27.5	-0.3	0.7955	
batchBatch_b080	-5.2	27.4	-0.2	0.8511	
batchBatch_b081	-3.4	27.5	-0.1	0.901	
batchBatch_b082	-0.9	27.4	0.0	0.9727	
batchBatch_b083	65.7	27.3	2.4	0.016	*
batchBatch_b084	55.3	27.4	2.0	0.0433	*
batchBatch_b085	-13.3	27.4	-0.5	0.6268	
batchBatch_b086	7.4	27.4	0.3	0.7877	
batchBatch_b087	10.2	27.5	0.4	0.7109	
batchBatch_b088	0.7	27.3	0.0	0.9782	
batchBatch_b089	-23.6	27.5	-0.9	0.3905	
batchBatch_b090	64.7	27.5	2.4	0.0185	*
batchBatch_b091	36.6	27.6	1.3	0.1849	
batchBatch_b092	20.0	27.5	0.7	0.4678	
batchBatch_b093	12.9	27.5	0.5	0.6385	
batchBatch_b094	-11.5	34.4	-0.3	0.7389	
batchBatch_b095	59.2	27.7	2.1	0.0326	*
batchUKBiLEVEAX_b1	14.7	26.8	0.5	0.5834	
batchUKBiLEVEAX_b10	-27.8	26.8	-1.0	0.2999	
batchUKBiLEVEAX_b11	-28.2	26.8	-1.1	0.2928	
batchUKBiLEVEAX_b2	-21.5	26.8	-0.8	0.4212	
batchUKBiLEVEAX_b3	-37.6	26.8	-1.4	0.1614	
batchUKBiLEVEAX_b4	-24.4	26.7	-0.9	0.3605	
batchUKBiLEVEAX_b5	6.4	26.8	0.2	0.8101	
batchUKBiLEVEAX_b6	-25.0	26.8	-0.9	0.3519	
batchUKBiLEVEAX_b7	-39.5	26.9	-1.5	0.142	
batchUKBiLEVEAX_b8	-13.2	26.8	-0.5	0.622	
batchUKBiLEVEAX_b9	NA	NA	NA	NA	
pc1	0.0	1.2	0.0	0.971	
pc2	-2.7	1.3	-2.1	0.0385	*
pc3	0.9	1.2	0.8	0.4505	
pc4	4.2	1.0	4.4	9.91E-06	***
pc5	1.1	0.4	2.6	0.0106	*
pc6	-2.5	1.2	-2.1	0.0349	*
pc7	0.2	1.1	0.2	0.8449	
pc8	-1.8	1.1	-1.7	0.0894	
pc9	-1.9	0.5	-3.9	8.98E-05	***

pc10	-2.5	1.0	-2.4	0.016	*
pc11	-2.0	0.8	-2.5	0.013	*
pc12	5.1	1.0	4.9	9.36E-07	***
pc13	0.9	1.2	0.7	0.4624	
pc14	-3.7	0.7	-5.6	1.85E-08	***
pc15	0.2	1.1	0.2	0.8629	
pc16	-3.4	0.7	-5.0	5.16E-07	***
pc17	0.4	1.0	0.4	0.6544	
pc18	0.4	0.7	0.6	0.5242	
pc19	-1.0	0.7	-1.5	0.1235	
pc20	4.8	0.7	7.1	1.15E-12	***

Table S2 | List of SNPs significantly associated with the iDXA phenotypes.

GWS significant signals (p-value<1.25E-08) were compiled across all phenotypes. These were sorted within 1Mb genomic windows and checked against all known associations under the umbrella term "obesity" on the GWAS catalog, in April of 2018. SNPs within published loci and SNPs which were associated with BMI or WHR with p-value<1E-12 were excluded. This led to a list of 1251 SNPs, the first 10 of which are shown here (sorted by p-value) and the rest of which can be found on https://doi.org/10.7488/ds/2918 for the rest of the GWAS signals, alongside column information and GWS associations with other traits.

rsid	*Gene*	*Pheno-type*	*chr*	*pos*	*a1*	*a0*	*n*	*freq1*	*beta1*	*se*	*p*	*Novel*	*Visual QC*	*Inde-pendent*	*Repli-cable*
rs62396185	HIST1H2BE	Pear fat	6	26180634	C	G	392537	0.257	-0.038	0.003	5.83E-49	no	no	no	no
rs537940107	KLF1429A	Leg FMR	7	130474916	T	C	392538	0.493	0.032	0.002	7.63E-45	no	yes	no	no
rs536819494	ADAMTSL3	Android LMI	15	84576633	T	A	392529	0.514	-0.029	0.002	1.68E-36	no	no	yes	no
rs143384	GDF5	TLI	20	34025756	G	A	392537	0.403	-0.027	0.002	4.35E-32	no	no	yes	no
rs575331485	HHIP	Gynoid fat mass	4	145731478	T	A	392524	0.406	0.027	0.002	1.11E-31	no	no	yes	no
rs1182199	GNA12	Pear fat	7	2862542	A	C	392537	0.305	-0.028	0.002	2.14E-30	no	no	yes	no
rs62271373	TSC22D2	Leg FMR	3	150066540	A	T	392538	0.060	-0.056	0.005	2.31E-30	no	no	yes	no
rs10786156	PLCE1	Android LMI	10	96014622	G	C	392529	0.433	0.026	0.002	5.49E-30	no	yes	yes	no
rs2135877	ADAMTSL3	Leg FMR	15	84663107	G	A	392538	0.327	0.027	0.002	9.51E-30	no	no	no	no

Table S3 | Systematic review & functional assessment of the 6 replicated iDXA GWAS signals.

Top SNPs at the 6 replicated loci were followed up using publicly available information to prioritise them in terms of their biological plausibility towards the phenotype, all accessed in July of 2018. rsid denotes the ID of the SNP. Genomic locations given in GRCh37. eQTL information via (Lonsdale et al., 2013), GWAS catalog via (Buniello et al., 2018), predicted effects on TFBSs via (Kumar et al., 2016), genes annotated on the basis of proximity within 1Mb windows from the lead SNP, gene function via (Stelzer et al., 2016), mouse phenotypes via (Dickinson et al., 2016) and mouse human protein identity given as a percentage of identical residues and via (Boratyn et al., 2012). Continues overleaf.

rsid	*Phenotype*	*snpid*	*a1*	*In gene*	*eQTL for*	*GWAS Catalog*	*Altered TFBS*
rs4820325	Leg FMR	22_38599978	A	MAFF	See Table S5	breast density, trigs, total fat (not GWS)	NA
rs1047891	Android LMI	2_211540507	A	CPS1	LANCL1	blood metabolites, BMI (not GWS)	NFIC
rs13072731	Android LMI	3_38533335	C	EXOG	EXOG & XYLB	P-wave duration	NA
rs17325374	Android LMI	2_230822858	G	FBXO36	TRIP12	NA	NA
rs12359330	Leg FMR	10_72414845	T	ADAMTS14	ADAMTS14	lymphocytes, extreme obesity (not GWS)	NA
rs2242449	Total Lean %	17_7095507	T	DLG4	ACADVL & ELP5	NA	ZNF263

rsid	*Genes*	*Gene function*	*Mouse KO*	*Mouse-Human Protein identity*
rs4820325	PLA2G6	phospholipase	infertile males	89.71%
	MAFF	transcription factor	NA	97.18%
	TMEM184B	heparin receptor & vesicle transport	shorter & enlarged heart	96.07%
rs1047891	CPS1	urea cycle	lethal	95.33%
	LANCL1	membrane protein	NA	91.48%
rs13072731	EXOG	single-stranded DNA and 5'-3' exonuclease	NA	83.15%
	XYLB	xylulokinase	various body compition	87.69%
rs17325374	FBXO36	ubiquitin-protein ligase	decreased blood cell counts	76.88%
	TRIP12	ubiquitin-protein ligase	NA	98.00%
rs12359330	ADAMTS14	aminoprocollagen peptidase activity	bone mineral density, cataract, abnormal vocalisation	81.55%
rs2242449	DLG4	synaptic channel subunit	decreased blood cell counts and various body composition	99.59%
	ACADVL	mitochondrial fatty acid beta-oxidation	NA	84.15%
	ELP5	RNA polymerase II elongator complex	NA	75.17%

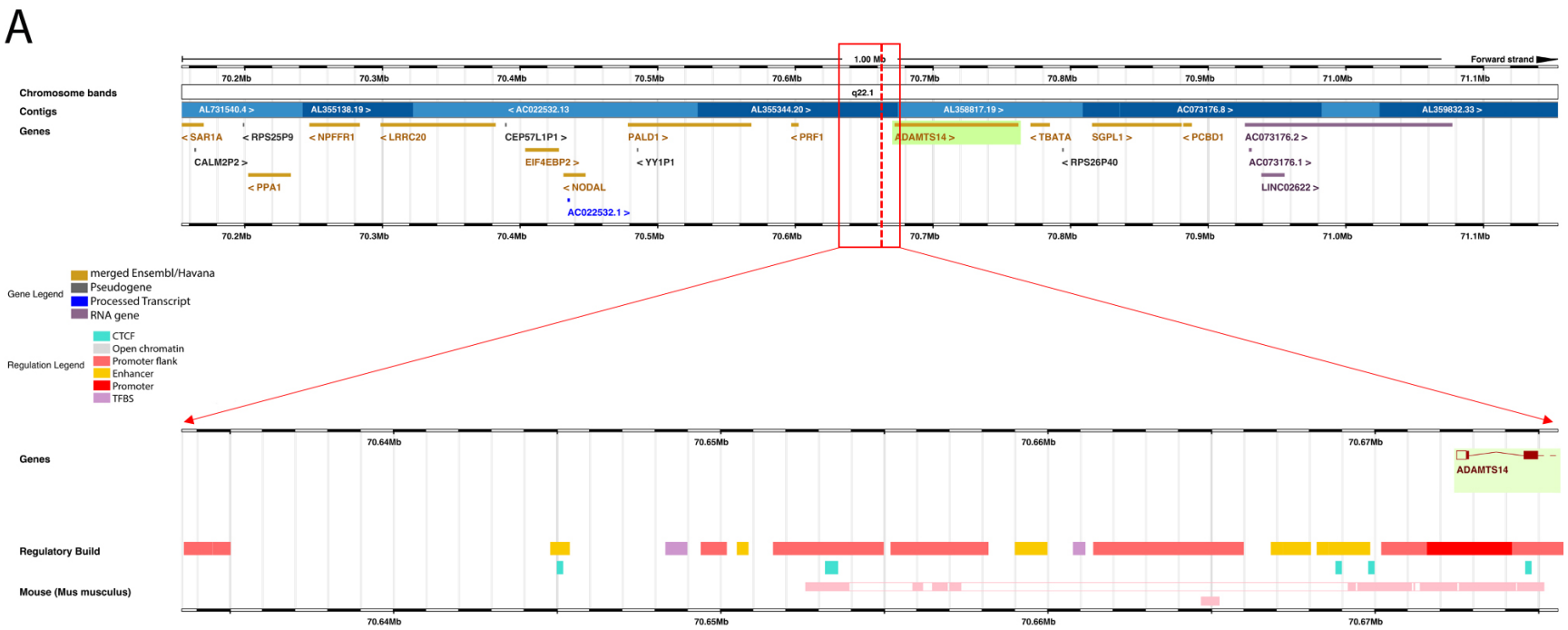
Table S4 | iDXA GWAS summary statistics for the lead variants at the PLA2G6 and ADAMTS14 loci.

Only GWS associations shown ($p<1.25x10^{-8}$). rsid denotes the ID of the SNP. Genomic locations given in GRCh37. Locus gene annotations made by proximity (+/- 500kb) to the lead associated variant using dbSNP build 150. a1 denotes the effect allele, with beta1 being the effect per copy of a1 on the phenotype, se being the standard error of beta1 and p being the p-value for this association. a0 is the other allele, n is the sample size and freq1 is the frequency of a1 in n.

Phenotype	*Locus*	*rsid*	*chr*	*pos*	*a1*	*a0*	*n*	*freq1*	*beta1*	*se*	*p*
Leg FMR	ADAMTS14	rs12359330	10	72414845	T	C	392538	0.2679	0.021148	0.002555	1.32E-16
Total Lean %	ADAMTS14	rs12359330	10	72414845	T	C	392532	0.2679	-0.017969	0.002555	2.07E-12
Leg FMR	PLA2G6	rs4820325	22	38599978	A	G	392538	0.5816	-0.022486	0.00229	9.61E-23
Pear fat	PLA2G6	rs4820325	22	38599978	A	G	392537	0.5816	-0.01731	0.00229	4.11E-14
Gynoid fat mass	PLA2G6	rs4820325	22	38599978	A	G	392524	0.5816	-0.015675	0.00229	7.72E-12
Total Lean %	PLA2G6	rs4820325	22	38599978	A	G	392532	0.5816	0.014978	0.00229	6.17E-11
Leg FMI	PLA2G6	rs4820325	22	38599978	A	G	392534	0.5816	-0.014609	0.00229	1.80E-10
SFI	PLA2G6	rs4820325	22	38599978	A	G	392535	0.5816	-0.013599	0.00229	2.90E-09

Figure S4 | Genomic context and sequence conservation at the ADAMTS14 (A) and PLA2G6 (B) loci.

Upper panels showing 1Mb surrounding the lead SNPs and association intervals. Areas highlighted in red indicate the LD blocks of the association signals and are expanded in the lower panels. Lower panels indicate the location of regulatory elements and mouse-human pairwise DNA sequence alignments within these blocks. Accessed via (Chen et al., 2010) in August of 2020.



B

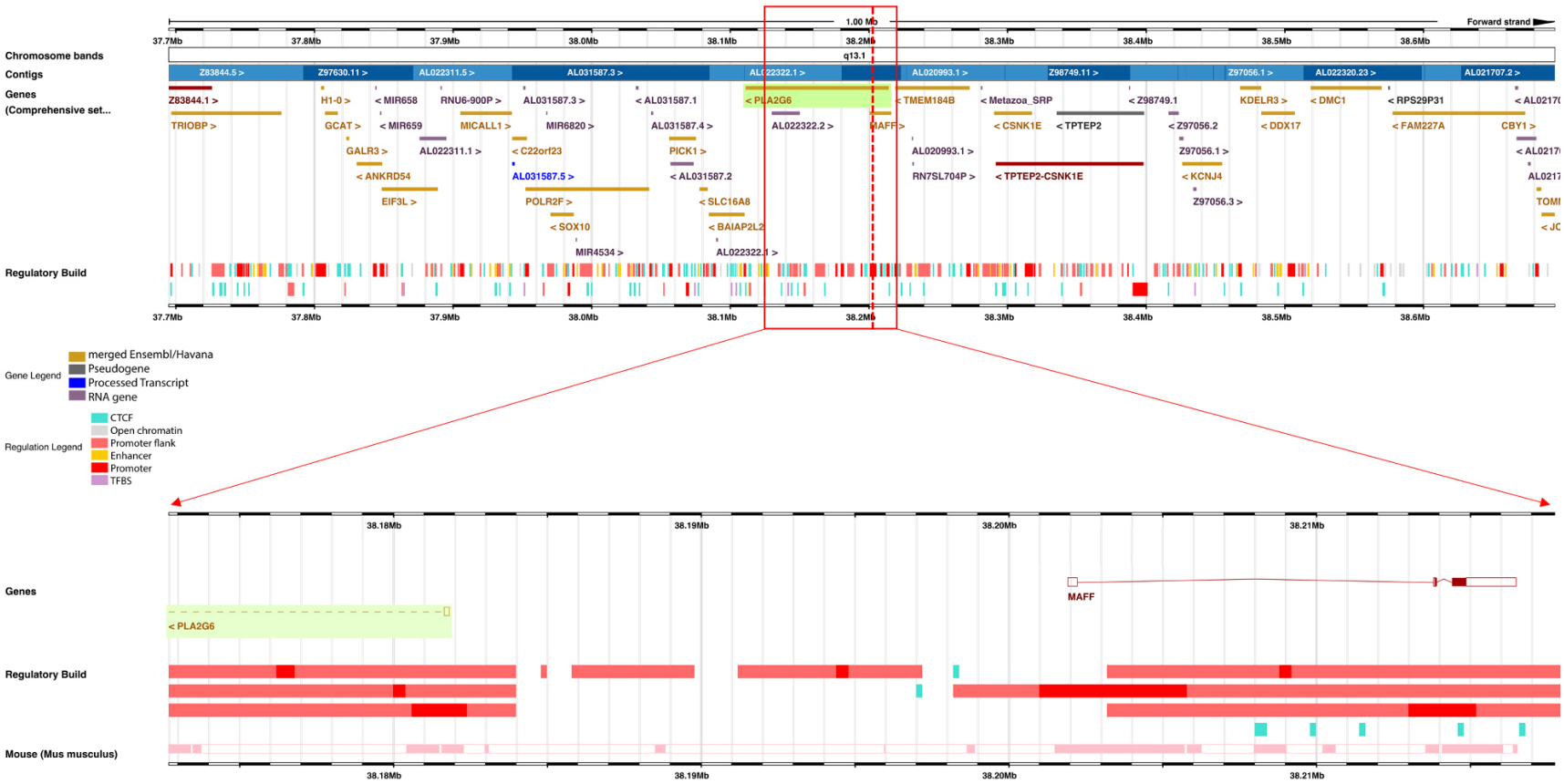

1.00 Mb
Forward strand
37.7Mb
37.8Mb
37.9Mb
38.0Mb
38.1Mb
38.2Mb
38.3Mb
38.4Mb
38.5Mb
38.6Mb
Chromosome bands
q13.1
Contigs
Z83844.5 >
Z97630.11 >
AL022311.5 >
AL031587.3 >
AL022322.1 >
AL020993.1 >
Z98749.11 >
Z97056.1 >
AL022320.23 >
AL021707.2 >
Genes
(Comprehensive set...
Z83844.1 >
H1-0 >
< MIR658
RNU6-900P >
AL031587.3 >
< AL031587.1
< PLA2G6
< TMEM184B
< Metazoa_SRP
< Z98749.1
KDELR3 >
< DMC1
< RPS29P31
TRIOBP >
GCAT >
< MIR659
MICALL1 >
MIR6820 >
AL031587.4 >
AL022322.2 >
MAFF >
< CSNK1E
< TPTEP2
< Z97056.2
< DDX17
< FAM227A
CBY1 >
GALR3 >
AL022311.1 >
< C22orf23
PICK1 >
AL020993.1 >
Z97056.1 >
< ANKRD54
AL031587.5 >
< AL031587.2
RN7SL704P >
< TPTEP2-CSNK1E
< KCNJ4
EIF3L >
POLR2F >
< SLC16A8
Z97056.3 >
< SOX10
< BAIAP2L2
MIR4534 >
AL022322.1 >
Regulatory Build
Gene Legend
merged Ensembl/Havana
Pseudogene
Processed Transcript
RNA gene
Regulation Legend
CTCF
Open chromatin
Promoter flank
Enhancer
Promoter
TFBS
38.18Mb
38.19Mb
38.20Mb
38.21Mb
Genes
MAFF
< PLA2G6
Regulatory Build
Mouse (Mus musculus)

Table S5 | Single-tissue eQTLs for rs4820325 at the *PLA2G6* locus implicate multiple genes.

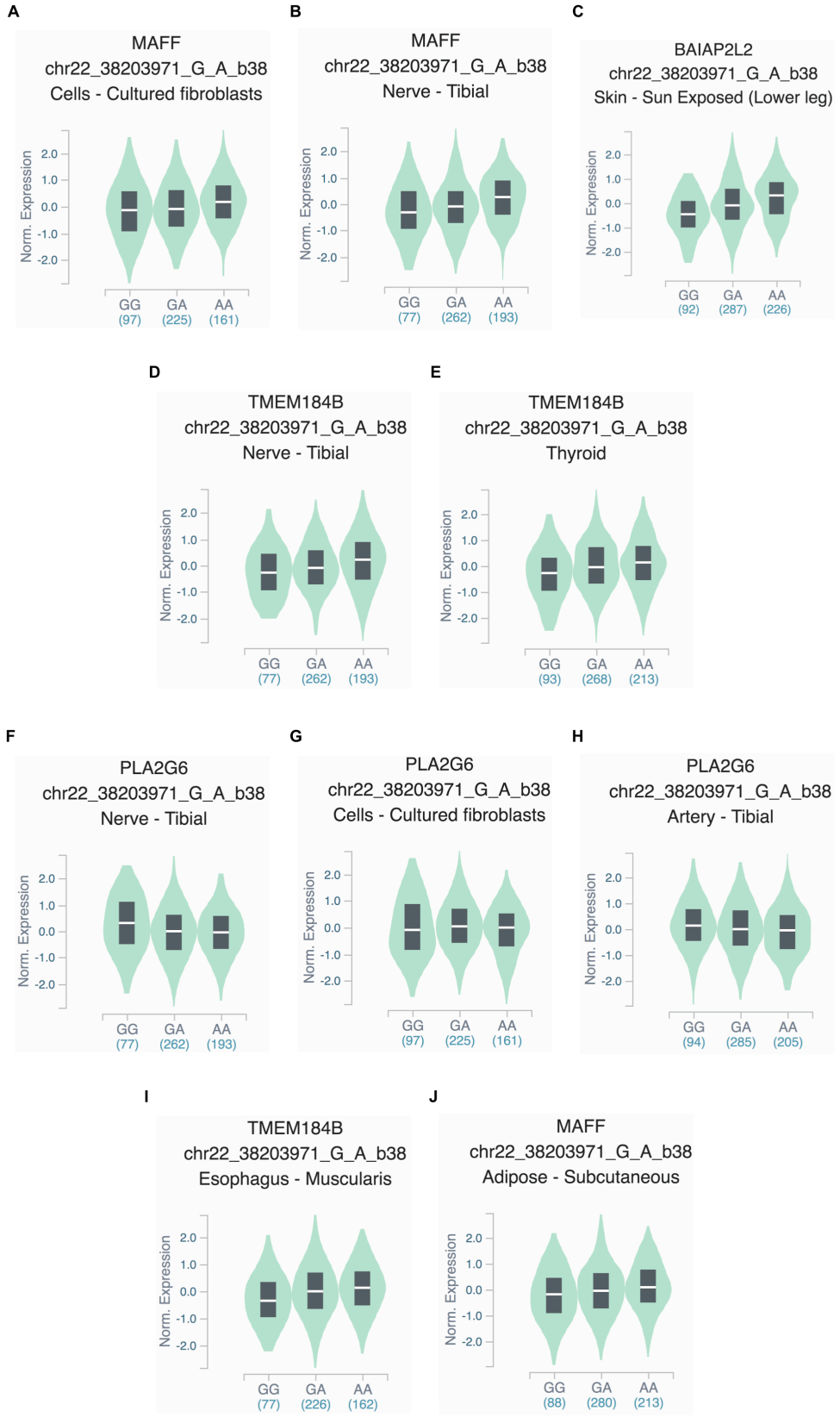
Direct eQTL data for the index SNP at the *PLA2G6* locus and all significant associated genes-tissue pairs, via https://gtexportal.org (as described in (Lonsdale et al., 2013) and accessed in April 2020 (Version 8)). For each gene, expression levels were normalised based on sample read counts per tissue and across samples using inverse-normal transformation. Displayed P-values were generated for each variant-gene pair, under the null hypothesis that the slope of a linear regression model between genotype and expression is zero. The displayed effect sizes represent the slope of that regression and were computed as the effect of the effect allele, relative to the reference allele in the human genome reference GRCh38/hg38. Results are sorted by increasing P-value.

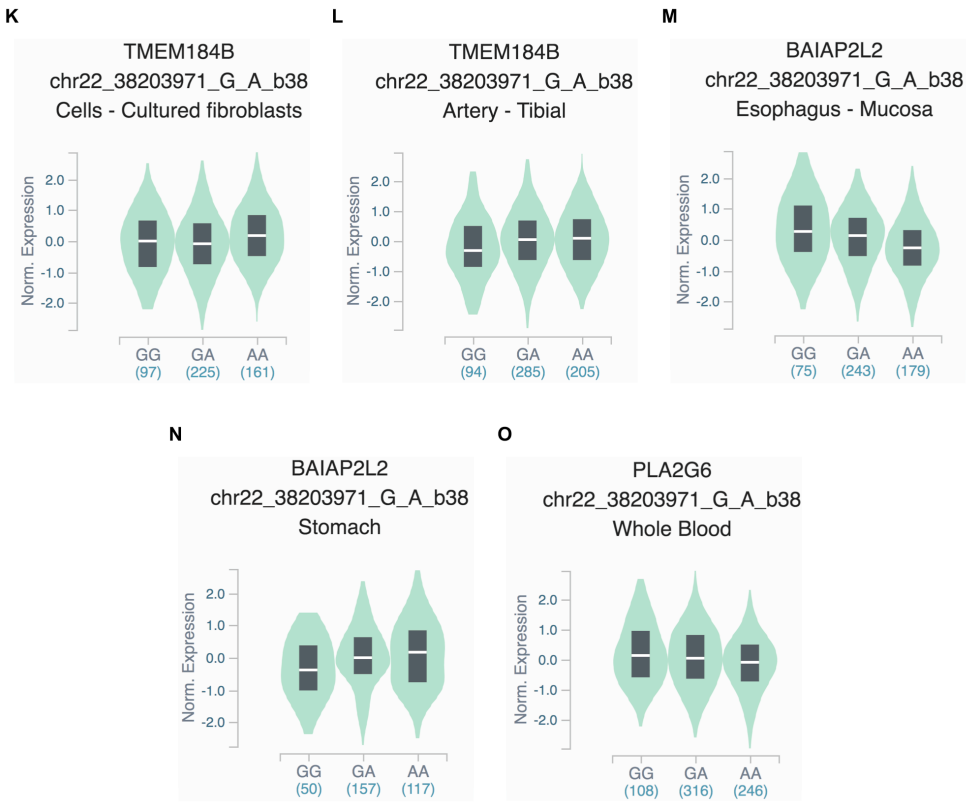
Gene	*SNP*	*P-value*	*Effect size*	*Tissue*
MAFF	rs4820325	1.40E-17	0.15	Cells - Cultured fibroblasts
MAFF	rs4820325	1.00E-16	0.2	Nerve - Tibial
BAIAP2L2	rs4820325	7.30E-15	0.3	Skin - Sun Exposed (Lower leg)
PLA2G6	rs4820325	6.30E-14	-0.18	Thyroid
TMEM184B	rs4820325	9.60E-13	0.22	Nerve - Tibial
TMEM184B	rs4820325	4.00E-12	0.2	Thyroid
PLA2G6	rs4820325	9.20E-11	-0.16	Nerve - Tibial
PLA2G6	rs4820325	1.70E-10	-0.087	Cells - Cultured fibroblasts
PLA2G6	rs4820325	3.60E-10	-0.16	Artery - Tibial
TMEM184B	rs4820325	4.50E-10	0.21	Esophagus - Muscularis

MAFF	rs4820325	4.50E-10	0.14	Adipose - Subcutaneous
TMEM184B	rs4820325	5.80E-10	0.15	Cells - Cultured fibroblasts
TMEM184B	rs4820325	5.80E-09	0.14	Artery - Tibial
BAIAP2L2	rs4820325	6.10E-09	-0.31	Esophagus - Mucosa
BAIAP2L2	rs4820325	5.90E-08	0.21	Stomach
PLA2G6	rs4820325	1.40E-07	-0.11	Whole Blood
MAFF	rs4820325	2.10E-07	0.14	Esophagus - Muscularis
PLA2G6	rs4820325	3.10E-07	-0.25	Pituitary
TMEM184B	rs4820325	3.20E-07	0.12	Adipose - Subcutaneous
BAIAP2L2	rs4820325	4.20E-07	0.21	Skin - Not Sun Exposed (Suprapubic)
TMEM184B	rs4820325	4.60E-07	0.13	Skin - Not Sun Exposed (Suprapubic)
TMEM184B	rs4820325	4.70E-07	0.13	Breast - Mammary Tissue
PLA2G6	rs4820325	6.20E-07	-0.11	Lung
TMEM184B	rs4820325	7.50E-07	0.19	Colon - Sigmoid
PLA2G6	rs4820325	1.00E-06	-0.096	Esophagus - Mucosa
TMEM184B	rs4820325	1.10E-06	0.13	Skin - Sun Exposed (Lower leg)
PLA2G6	rs4820325	1.30E-06	-0.16	Heart - Atrial Appendage
TMEM184B	rs4820325	1.30E-06	0.16	Heart - Atrial Appendage

PLA2G6	rs4820325	2.00E-06	-0.13	Artery - Aorta
PLA2G6	rs4820325	2.70E-06	-0.11	Esophagus - Muscularis
SLC16A8	rs4820325	3.50E-06	0.15	Artery - Aorta
BAIAP2L2	rs4820325	7.20E-06	0.16	Adipose - Subcutaneous
MAFF	rs4820325	7.50E-06	0.1	Artery - Tibial
PLA2G6	rs4820325	1.10E-05	-0.16	Pancreas
PLA2G6	rs4820325	1.20E-05	-0.18	Spleen
PLA2G6	rs4820325	1.60E-05	-0.23	Brain - Caudate (basal ganglia)
MAFF	rs4820325	2.10E-05	0.2	Artery - Coronary
PLA2G6	rs4820325	2.70E-05	-0.14	Esophagus - Gastroesophageal Junction
TMEM184B	rs4820325	5.30E-05	0.11	Adipose - Visceral (Omentum)
TMEM184B	rs4820325	5.40E-05	0.099	Lung
PLA2G6	rs4820325	5.70E-05	-0.16	Stomach
BAIAP2L2	rs4820325	8.80E-05	-0.095	Cells - Cultured fibroblasts
TMEM184B	rs4820325	9.00E-05	0.13	Artery - Aorta

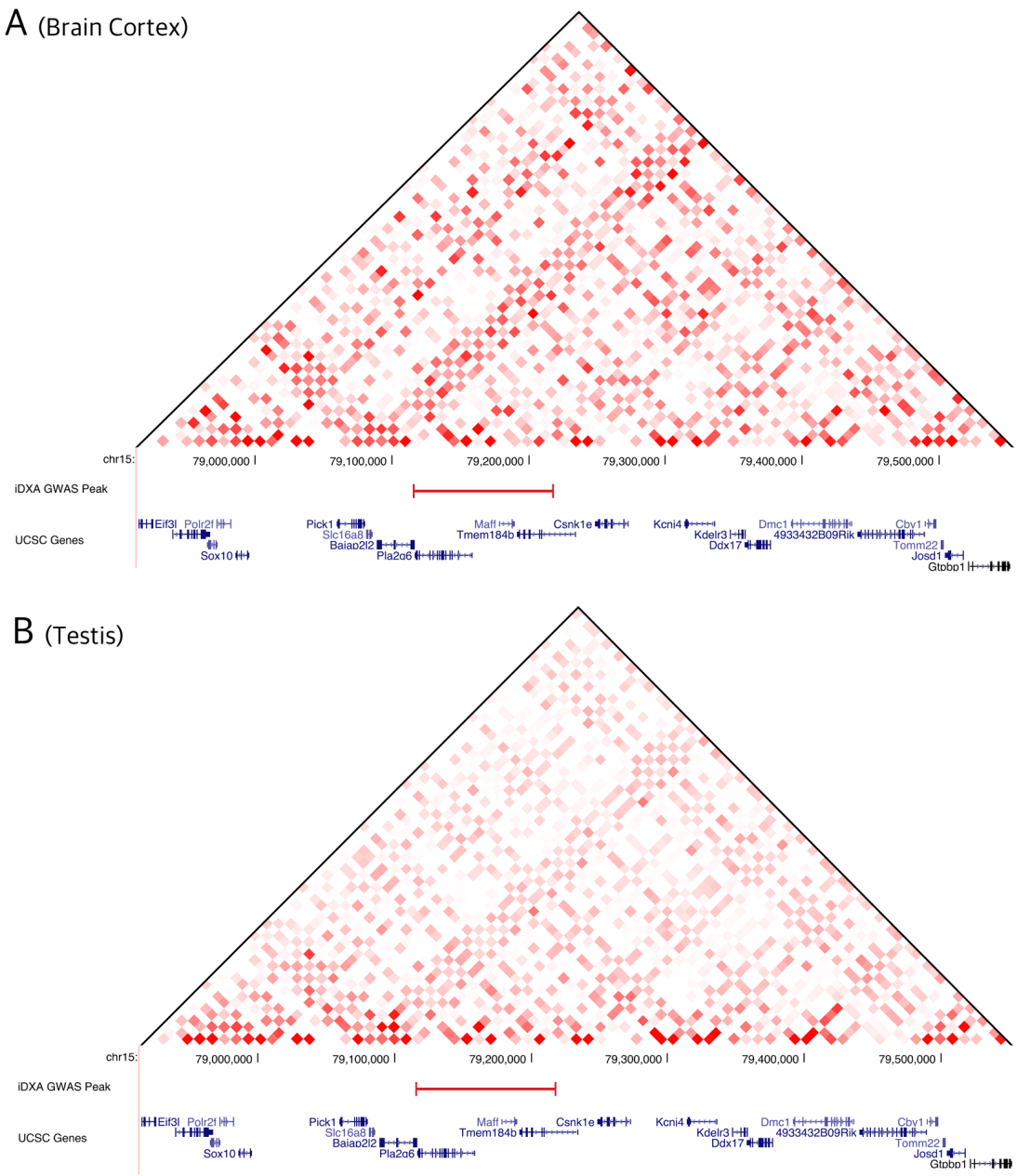
Figure S5 | Top 15 single-tissue eQTLs for rs4820325 at the *PLA2G6* locus.





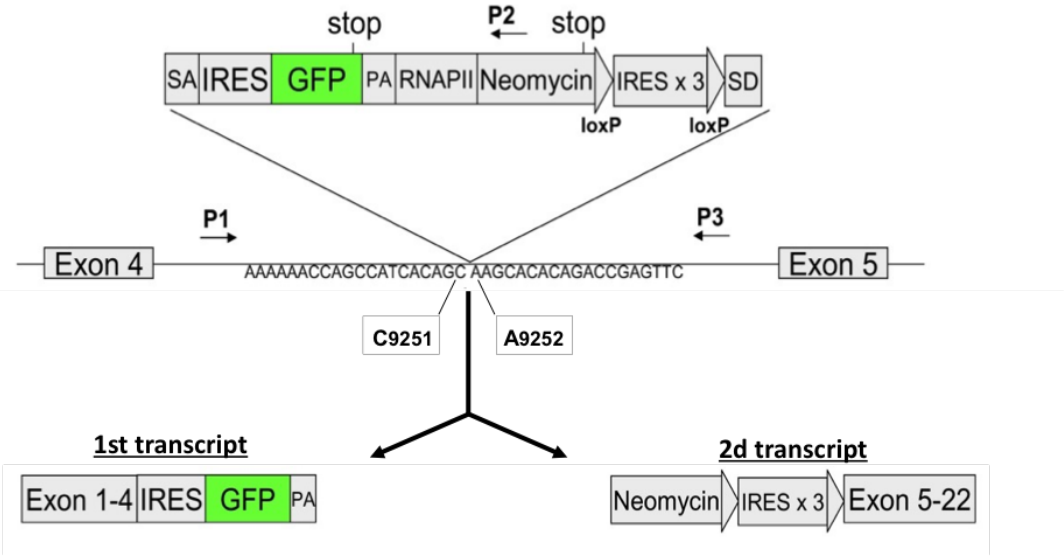
The 15 most significant single tissue eQTLs from Table S1, displayed as differences in expression levels against genotype, for SNP rs4820325. Data taken from https://gtexportal.org (Lonsdale et al., 2013).

Figure S6 | Normalised, non-binned heatmaps of the 5C data.



Chromatin interactions within the 650kb window on chromosome 15 in mouse-derived brain cortex cells (A) and testis cells (B). Interaction frequencies were normalized to an unchanging gene desert region on chromosome 15. GWAS peak area indicated in red.

Figure S7 | Generation of the *Adamts14*$^{-/-}$ mouse.



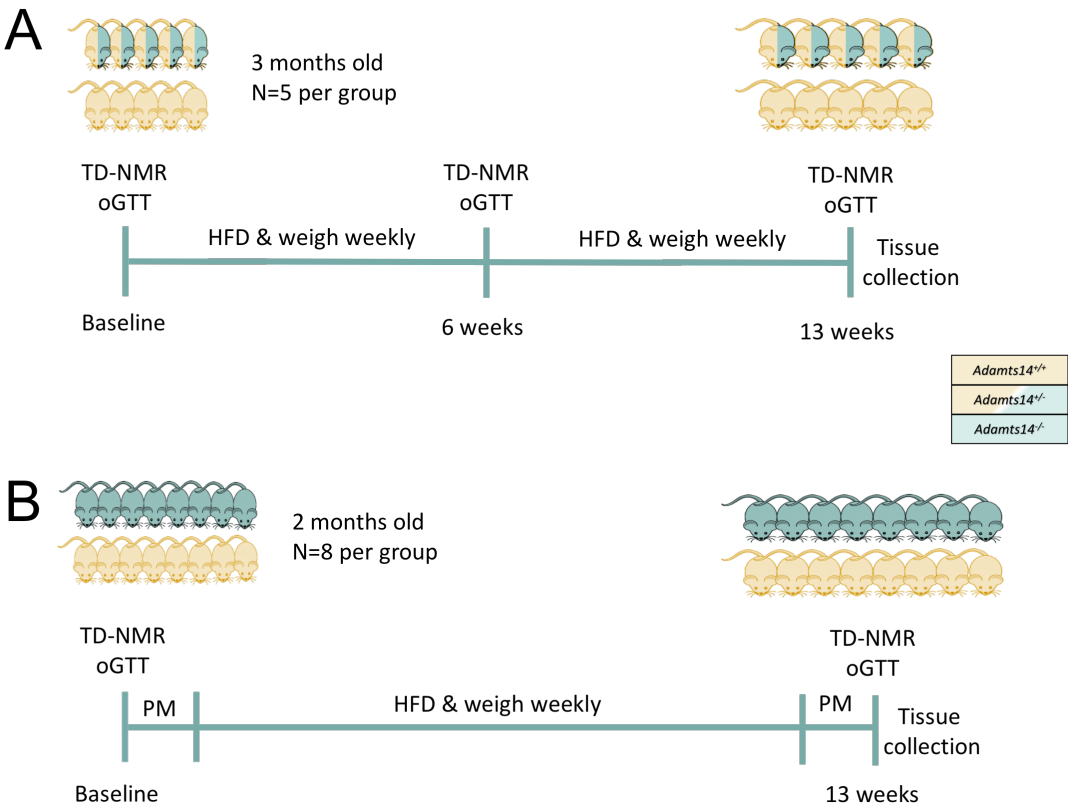
The *Adamts14*$^{-/-}$ mouse was created as outlined in (Dupont et al. 2018) and kindly given to us by the Colige Group, from the University of Liege. Briefly, the genetrap was inserted in intron 4 of the gene and was made up of several elements, including a Green Fluorescent Protein (GFP) sequence. This insertion caused the expression of two transcripts, one under the *Adamts14* promoter and one under an RNA polymerase II promoter. The indicated primers (P1, P2 and P3) were used by Dupont et al. for genotyping. *Taken from the supplement in (Dupont et al., 2018).*

Table S6 | Primers used for the genotyping of the *Adamts14*$^{-/-}$ mice.

The WT reaction from Dupont et al. (P1&P3) worked fine and yielded the expected amplicon size, for the KO reaction (P1&P2) the amplicon size or KO sequence had not been reported. Thus, I created a GFP reaction to differentiate between heterozygous and WT animals. To later distinguish the homozygous animals, I designed genotyping reactions that spanned the break site, where the genetrap cassette had inserted. These two assays were run as follows: 95°C for 5 minutes, 30 cycles of 95°C for 30 seconds, 65°C for 45 seconds and 72°C for 45 seconds, then 72°C for 5 minutes and according to the Platinum™ SuperFi™ PCR Master Mix (Thermofisher, 12358250) suggested concentrations.

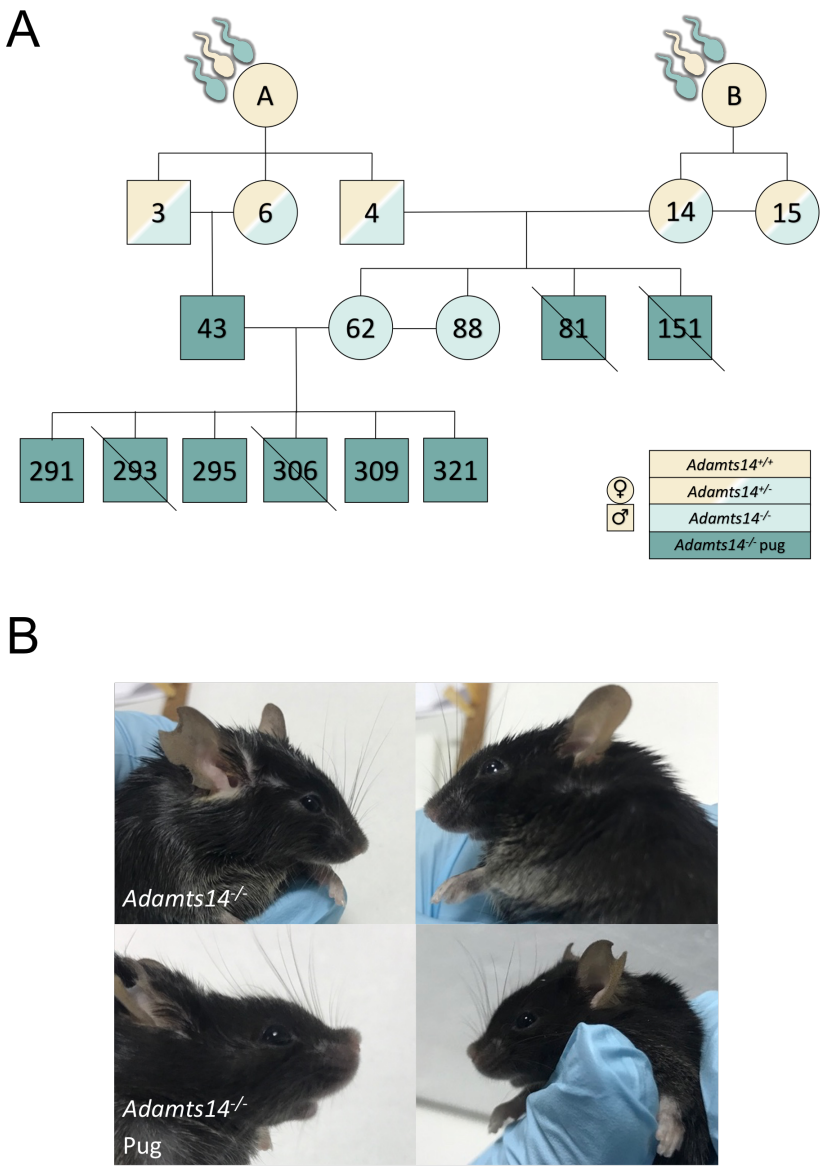
Name	*Location*	*With*	*Sequence (5' to 3')*	*Amplicon*
P1	Intron 4, upstream of insertion site	P2 or P3	TTAATGGCTCACTCCCCACTGAAC	-
P2	Within cassette	P1	TTCCTTGGGAGGGTCTCCTCTGA	(?) bp
P3	Intron 4, downstream of the insertion site	P1	GCCAGGAAGGACCAGTGATCCTC	354 bp
F_{GFP}	Within the GFP domain of the cassette	R_{GFP}	AGCTGGACGGCGACGTAAAC	-
R_{GFP}	As above	F_{GFP}	GCGCTTCTCGTTGGGGTCTT	596 bp
F_{WT}	Downstream of insertion site	R_{KO} or R_{KO}	GTTCAGTGGGGAGTGAGCCATTAA	-
R_{WT}	Spans insertion site	F_{WT}	CTGTGTGCTTGCTGTGATGGCTG	252 bp
R_{KO}	Spans insertion site	F_{WT}	CCTGGACCAGCTGTGATGGCTG	251 bp

Figure S8 | Experimental design for the phenotyping of the *Adamts14*$^{+/-}$ and *Adamts14*$^{-/-}$mice.



Schematic representation of the experimental design for the *in vivo* metabolic characterisation of the *Adamts14*$^{+/-}$ (A) and *Adamts14*$^{-/-}$ (B) mice. Both cohorts were exposed to HFD and weighed weekly for 13 weeks, with glucose homeostasis and body composition being assessed at the start and end of the HFD exposure, via oGTT and TD-NMR respectively. Heterozygous animals underwent both assessments at the midpoint of the experiment as well (6 weeks). Homozygous animals underwent indirect calorimetry (indicated by PM) at the start and end of the experiment. Tissues were collected at the end of the experiments for further analysis.

Figure S9 | Some male *Adamts14*$^{-/-}$ mice exhibit airorhynchic brachycephaly.



(A) Sperm from *Adamts14*$^{+/-}$ C57Bl6/J animals was received from the Colige group and was used to perform IVF on WT females. The colony was expanded to obtain *Adamts14*$^{-/-}$ animals. Some of the resulting *Adamts14*$^{-/-}$ male mice developed brachycephaly with an upwards nasal tilt. This phenotype is similar to dog breeds such as pugs, but the nasal tilt has not been seen in any other brachycephalic mice (Jeffrey Schoenebeck, pers. comm.) (B) Comparison of two *Adamts14*$^{-/-}$ mice, littermates 150 and 151, the latter of which was affected. These mice behaved and bred normally, but were apparently smaller than their male and female littermates.